LARGE-SCALE MINIMIZATION OF THE PSEUDOSPECTRAL ABSCISSA

NICAT ALIYEV* AND EMRE MENGI†

Abstract. This work concerns the minimization of the pseudospectral abscissa of a matrix-valued function dependent on parameters analytically. The problem is motivated by robust stability and transient behavior considerations for a linear control system that has optimization parameters. We describe a subspace procedure to cope with the setting when the matrix-valued function is of large size. The proposed subspace procedure solves a sequence of reduced problems obtained by restricting the matrix-valued function to small subspaces, whose dimensions increase gradually. It possesses desirable features such as the global convergence of the minimal values of the reduced problems to the minimal value of the original problem, and a superlinear convergence exhibited by the decay in the errors of the minimizers of the reduced problems. In mathematical terms, the problem we consider is a large-scale nonconvex minimax eigenvalue optimization problem such that the eigenvalue function appears in the constraint of the inner maximization problem. Devising and analyzing a subspace framework for the minimax eigenvalue optimization problem at hand with the eigenvalue function in the constraint require special treatment that makes use of a Lagrangian and dual variables. There are notable advantages in minimizing the pseudospectral abscissa over maximizing the distance to instability or minimizing the $\mathcal{H}_\infty$ norm; the optimized pseudospectral abscissa provides quantitative information about the worst-case transient growth, and the initial guesses for the parameter values to optimize the pseudospectral abscissa can be arbitrary, unlike the case to optimize the distance to instability and $\mathcal{H}_\infty$ norm that would normally require initial guesses yielding asymptotically stable systems.

Key words. pseudospectral abscissa, large scale, subspace framework, Lagrangian, robust stability, eigenvalue optimization, nonconvex optimization

AMS subject classifications. 65F15, 93C05, 93D09, 90C26, 90C47

1. Introduction.

The minimization of the spectral abscissa of a linear control system has drawn interest in the last couple of decades [15, 4]. A classical problem that can be tackled using the spectral abscissa minimization is the stabilization by static output feedback (SOF) problem; given matrices $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, find a controller $K \in \mathbb{C}^{m\times p}$ such that the system $x'(t) = (A + BKC)x(t)$ is asymptotically stable, equivalently $A + BKC$ has all of its eigenvalues on the open left-half of the complex plane. SOF is known to be a notoriously difficult problem [9]. Indeed, it has been shown that SOF when the entries of K are subject to box constraints is NP-hard [10, 34]. Mathematically, the spectral abscissa minimization is a nonconvex eigenvalue optimization problem that involves the minimization of the real part of the rightmost eigenvalue. Nonsmoothness at a locally optimal point minimizing the spectral abscissa can occur due to possible nonsimplicity of the rightmost eigenvalue at the local optimizer [13, 14], as well as due to the existence of multiple rightmost eigenvalues with the same real part; see, e.g., the Turbo generator example in [16]. Numerically speaking a bigger challenge is the non-Lipschitz nature of the rightmost eigenvalue at a local optimizer; when the rightmost eigenvalue is not simple at a local optimizer, it can change rapidly near the optimizer. In more formal terms, the spectral abscissa does not have to be Lipschitz continuous, not even locally, at a local optimizer, and this causes numerical difficulties to numerical algorithms, which

*Czech Technical University, Department of Instrumentation and Control, Technicka 4, 16607, Prague, Czech Republic (nijat.aliyev@cvut.cz).

†Koç University, Department of Mathematics, Rumeli Feneri Yolu 34450, Sarıyer, Istanbul, Turkey (emengi@ku.edu.tr).

are at least prone to rounding errors. A second difficulty with minimizing the spectral abscissa is that, even if a negative spectral abscissa guarantees an asymptotic decay, the system can still exhibit significant transient growth before the eventual decay.

As a remedy to these problems with the spectral abscissa minimization, the pseudospectral abscissa minimization has been considered in the last two decades [16, 3]. Recall that the spectral abscissa of a matrix $A \in \mathbb{C}^{n\times n}$ is given by $\alpha(A) := \max\{\mathrm{Re}(z) \mid z \in \Lambda(A)\}$, where $\Lambda(\cdot)$ denotes the spectrum (i.e., the set of all eigenvalues) of its matrix argument. On the other hand, for a given real number $\epsilon > 0$, the *ϵ-pseudospectrum* of A, which we denote by $\Lambda_\epsilon(A)$, consists of eigenvalues of all matrices within an ϵ-neighborhood of A, formally defined as

$$\Lambda_\epsilon(A) \;:=\; \left\{ z \in \mathbb{C} \mid z \in \Lambda(A+\Delta) \;\; \exists \Delta \in \mathbb{C}^{n\times n} \;\; \text{s.t.} \;\; \|\Delta\|_2 \le \epsilon \right\},$$

and the *ϵ-pseudospectral abscissa* of A is defined as $\alpha_\epsilon(A) := \max\{\mathrm{Re}(z) \mid z \in \Lambda_\epsilon(A)\}$, that is as the real part of the rightmost point in $\Lambda_\epsilon(A)$ [37, 38].

Unlike the spectral abscissa, the pseudospectral map $A \mapsto \alpha_\epsilon(A)$ is locally Lipschitz continuous [30, Corollary 7.2], [22, Corollary 3.4]. Moreover, if $\alpha_\epsilon(A)$ is negative, the system at hand as well as all nearby systems at a distance of ϵ are asymptotically stable. This is significant especially if the entries of the matrix are uncertain. A second virtue in using the pseudospectral abscissa rather than the spectral abscissa is that, if it is sufficiently small, not only the asymptotic decay but also a nice transient behavior of the solution of the autonomous system $x'(t) = Ax(t)$ may be possible, thanks to the Kreiss-matrix theorem [38, 25]. Equivalent and computationally plausible characterizations for $\Lambda_\epsilon(A)$ and $\alpha_\epsilon(A)$ are given by [38]

$$\begin{aligned} \Lambda_\epsilon(A) \;&=\; \{z \in \mathbb{C} \mid \sigma_{\min}(A - zI) \le \epsilon\}, \\ \alpha_\epsilon(A) \;&=\; \max\{\mathrm{Re}(z) \mid \sigma_{\min}(A - zI) \le \epsilon\}, \end{aligned} \tag{1.1}$$

where $\sigma_{\min}(\cdot)$ denotes the smallest singular value of its matrix argument. We remark that $\alpha_\epsilon(A)$, just like the spectral abscissa, is usually a nonsmooth function of the entries of A when the rightmost point in $\Lambda_\epsilon(A)$ is not unique, or the singular value $\sigma_{\min}(A - \widetilde{z}I)$ is not simple at a rightmost point $\widetilde{z}$ of $\Lambda_\epsilon(A)$.

The computation of the ϵ-pseudospectral abscissa and its derivatives require more work compared to that for the spectral abscissa, yet there are very good algorithms to compute the ϵ-pseudospectral abscissa. The criss-cross algorithm developed by Burke, Lewis and Overton is globally convergent at a quadratic rate, hence computes the pseudospectral abscissa very reliably and efficiently for small- to medium-size matrices [11, 17]. For the pseudospectral abscissa of larger-size matrices, there is a fixed-point iteration developed by Guglielmi et al. [20], which can further be accelerated with the subspace framework in [27]. We also refer to [7] for improvements on the quadratically convergent criss-cross algorithm, and its extension to spectral value sets.

1.1. Problem and Contributions. In this work, we assume we are given a matrix-valued function $A : \Omega \to \mathbb{C}^{n\times n}$ of the form

$$A(x) = f_1(x)A_1 + \cdots + f_\kappa(x)A_\kappa \tag{1.2}$$

dependent on the parameters x, where $A_1, \ldots, A_\kappa \in \mathbb{C}^{n\times n}$, the size of the matrices n is very large, and $f_1, \ldots, f_\kappa : \Omega \to \mathbb{R}$ are real-analytic on Ω, a nonempty, open subset of $\mathbb{R}^d$ representing the permissible values for the parameters. We deal with the ϵ-pseudospectral abscissa minimization problem, for a prescribed $\epsilon > 0$, which is the

minimax problem

$$\begin{aligned} \min_{x\in\underline{\Omega}} \alpha_\epsilon(A(x)) &= \min_{x\in\underline{\Omega}} \max\{\mathrm{Re}(z) \mid z\in\Lambda_\epsilon(A(x))\} \\ &= \min_{x\in\underline{\Omega}} \max\{\mathrm{Re}(z) \mid \sigma_{\min}(A(x)-zI)\le\epsilon\} \end{aligned} \tag{1.3}$$

over a compact, nonempty subset $\underline{\Omega}$ of Ω. We describe a subspace procedure that reduces the size of A considerably but without altering the optimal parameter values. We prove in theory that the subspace procedure possesses desirable convergence properties such as global convergence and a superlinear rate of convergence. As a result, it enables us to solve the ϵ-pseudospectral abscissa minimization problems involving parameter dependent matrices with sizes on the order of thousands.

The stabilization by static output feedback problem for given $A\in\mathbb{C}^{n\times n}$ $B\in\mathbb{C}^{n\times m}$, $C\in\mathbb{C}^{p\times n}$ can be treated in a robust way by minimizing $\alpha_\epsilon(A+BKC)$ over $K\in\mathbb{C}^{m\times p}$ with entries constrained to lie in prescribed intervals. If the minimal value of $\alpha_\epsilon(A+BKC)$ is negative, not only the system but also nearby systems are stabilizable. Such a stabilization problem falls into the setting of (1.3), as $A+BKC$ can represented in the form (1.2). Variants of Newton's Method, especially BFGS, with proper line-search have been successfully applied to such nonsmooth problems recently. The difficulty is that these smooth optimization techniques converge at a linear rate at best on nonsmooth problems. They would require the computation of $\alpha_\epsilon(A+BKC)$ in the objective quite a few times, which makes them prohibitively expensive in the large-scale setting. The package HANSO [35] is based on a hybrid method that makes use of both BFGS and a gradient sampling algorithm [15]. A specialization of this hybrid method for $\mathcal{H}_\infty$ controller design called HIFOO [12, 21, 6] is applicable for the minimization of the spectral abscissa and pseudospectral abscissa, as well as maximizing the distance to instability of $A+BKC$ over K. The more recent package GRANSO [18] is also based on BFGS, and is a variant of HANSO that can cope with box constraints. The downsides of all of these methods are that they converge locally, and they are not well-suited for large-scale problems. Locally-convergent bundling techniques and spectral bundle methods [3, 5, 4] are also used for solving problems related to SOF.

The problem at hand is a nonconvex large-scale minimax eigenvalue optimization problem. What makes it peculiar compared to our previous works [32, 2] is that the eigenvalue function appears in the constraint of the inner maximization problem. This is in contrast to [32] and [2] that introduce subspace frameworks for large-scale minimax eigenvalue optimization problems - specifically for the maximization of the distance to instability and minimization of the $\mathcal{H}_\infty$-norm, respectively - where the eigenvalue functions appear in the objective. Designing a subspace framework for a minimax problem with the eigenvalue function in the constraint, and especially analyzing its convergence require a special treatment. For instance, when establishing the global convergence of the proposed framework and its superlinear convergence, we work on the Lagrangian as well as the dual variable as much as the primal variables.

Minimizing the ϵ-pseudospectral abscissa of $A(x)$, that is the problem in (1.3), and maximizing the distance to instability of $A(x)$ (more generally minimizing the $\mathcal{H}_\infty$-norm for a linear time invariant system depending on parameters) are motivated by similar robust stability and transient behavior considerations. However, there are advantages in optimizing the pseudospectral abscissa. First, by minimizing the pseudospectral abscissa, we simultaneously minimize a concrete lower bound on the largest transient growth possible. Secondly, when maximizing the distance to instability or minimizing the $\mathcal{H}_\infty$ norm, the system with the initial guess for the parameter

values should ideally be asymptotically stable, and finding such a guess may be a challenge. When minimizing the pseudospectral abscissa, it does not matter to start with parameter values leading to systems not asymptotically stable.

1.2. Outline. We introduce the subspace framework in the next section, then investigate its properties such as the derivatives of the original and reduced problems, and, consequently, deduce Hermite interpolation properties between the original and reduced problems. The proposed framework produces a sequence of reduced problems with sizes increasing gradually. In Section 3, we formally show that if, in finite dimension, the minimizers of the reduced problems stagnate, then the point of stagnation is actually a global minimizer of the original problem, and, in infinite dimension, the globally minimal values of the reduced problems in the limit converge to the globally minimal value of the original problem. Section 4 is devoted to a rate-of-convergence analysis of the proposed framework. In this section, under mild assumptions, we prove a superlinear convergence result for the errors of the reduced problems. Section 5 discusses the extensions of the proposed framework to minimize the real ϵ-pseudospectral abscissa, the real part of the rightmost point in the real ϵ-pseudospectrum when the perturbations are constrained to be real matrices. A Matlab implementation of the proposed framework is made publicly available. In Section 6, we perform numerical experiments with this implementation on synthetic as well as benchmark examples from the *COMPl$_e$ib* collection [29], and observe that the deduced theoretical global convergence and superlinear convergence results hold in practice. The numerical results illustrate the efficiency and accuracy of the subspace framework in practice on large-scale pseudospectral abscissa minimization problems.

2. Subspace Framework. In this section, we present a subspace framework for the minimization of the pseudospectral abscissa of a large-scale parameter dependent matrix $A(x)$ of the form (1.2). We resort to one-sided projections to deal with the large size of $A(x)$. Specifically, given a subspace $\mathcal{V} \subseteq \mathbb{C}^n$ of dimension $k \ll n$, and a matrix $V \in \mathbb{C}^{n\times k}$ whose columns form an orthonormal basis for $\mathcal{V}$, we minimize the pseudospectral abscissa of the reduced matrix-valued function

$$A^V(x) := A(x)V = f_1(x)A_1V + \cdots + f_\kappa(x)A_\kappa V \tag{2.1}$$

instead of minimizing the pseudospectral abscissa of $A(x)$. Formally, defining the ϵ-pseudospectrum of the reduced matrix-valued function by

$$\Lambda_\epsilon(A^V(x)) \;:=\; \left\{z \in \mathbb{C} \;\middle|\; \sigma_{\min}(A^V(x) - zV) \le \epsilon\right\}$$

and the corresponding ϵ-pseudospectral abscissa by

$$\begin{aligned} \alpha_\epsilon(A^V(x)) \;&:=\; \max\left\{\mathrm{Re}(z) \;\middle|\; z \in \Lambda_\epsilon(A^V(x))\right\} \\ &=\; \max\left\{\mathrm{Re}(z) \;\middle|\; z \in \mathbb{C} \;\text{ s.t. }\; \sigma_{\min}(A^V(x) - zV) \le \epsilon\right\}, \end{aligned} \tag{2.2}$$

we solve

$$\min_{x\in\underline{\Omega}} \alpha_\epsilon(A^V(x)) \;=\; \min_{x\in\underline{\Omega}} \max\left\{\mathrm{Re}(z) \;\middle|\; z \in \mathbb{C} \text{ s.t. } \sigma_{\min}(A^V(x) - zV) \le \epsilon\right\} \tag{2.3}$$

rather than (1.3). Employing two-sided projections for the reduction of $A(x)$ may appear as a plausible strategy since $A(x)$ is a non-Hermitian matrix-valued function. Yet, two-sided projections cause convergence problems, such as the loss of a monotonicity property, crucial for the convergence analysis discussed in Section 3.

We remark that $\Lambda_\epsilon(A^V(x))$ and $\alpha_\epsilon(A^V(x))$ are independent of the choice of the orthonormal basis used for $\mathcal{V}$, that is $\Lambda_\epsilon(A^{V_1}(x)) = \Lambda_\epsilon(A^{V_2}(x))$ and $\alpha_\epsilon(A^{V_1}(x)) = \alpha_\epsilon(A^{V_2}(x))$ for two different matrices V_1, V_2 whose columns form orthonormal bases for $\mathcal{V}$, as $\sigma_{\min}(A^{V_1}(x) - zV_1) = \sigma_{\min}(A^{V_2}(x) - zV_2)$ for any $z \in \mathbb{C}$ for such V_1, V_2. Accordingly, letting $\mathcal{V} := \mathrm{Col}(V)$, and hiding the dependence of the pseudospectra and pseudospectral abscissa on $A(x)$, we use the shorthand notations

$$\Lambda_\epsilon(x) := \Lambda_\epsilon(A(x)), \quad \Lambda_\epsilon^{\mathcal{V}}(x) := \Lambda_\epsilon(A^V(x)),$$
$$\alpha_\epsilon(x) := \alpha_\epsilon(A(x)), \quad \alpha_\epsilon^{\mathcal{V}}(x) := \alpha_\epsilon(A^V(x))$$

throughout the rest of this text.

The basic subspace framework for the minimization of $\alpha_\epsilon(x)$ is described in Algorithm 2.1. At each subspace iteration in Algorithm 2.1 in lines 6–10, first a small-scale

Algorithm 2.1 The subspace framework to minimize $\alpha_\epsilon(x)$ over $\underline{\Omega}$

Input: The matrix-valued function $A(x)$, the feasible region $\underline{\Omega}$, and $\epsilon > 0$.

Output: An estimate $\widehat{x}$ for $\arg\min_{x\in\underline{\Omega}} \alpha_\epsilon(x)$, and $\widehat{z} \in \mathbb{C}$ that is an estimate for a globally rightmost point in $\Lambda_\epsilon(\widehat{x})$

1: $x_1^{(1)}, \ldots x_\eta^{(1)} \leftarrow$ initially chosen points in $\underline{\Omega}$.
2: $z_j^{(1)} \leftarrow \arg\max \left\{ \mathrm{Re}(z) \mid \sigma_{\min}(A(x_j^{(1)}) - zI) \le \epsilon \right\}$ for $j = 1, \ldots, \eta$.
3: $v_j^{(1)} \leftarrow$ a right singular vector corr. to $\sigma_{\min}(A(x_j^{(1)}) - z_j^{(1)} I)$ for $j = 1, \ldots, \eta$.
4: $\mathcal{V}_1 \leftarrow \mathrm{span}\left\{ v_1^{(1)}, \ldots v_\eta^{(1)} \right\}$ and $V_1 \leftarrow$ an orthonormal basis for $\mathcal{V}_1$.
5: **for** $k = 2, 3, \ldots$ **do**
6: $\quad x^{(k)} \leftarrow \arg\min_{x\in\underline{\Omega}} \alpha_\epsilon^{\mathcal{V}_{k-1}}(x)$.
7: $\quad z^{(k)} \leftarrow \arg\max \left\{ \mathrm{Re}(z) \mid \sigma_{\min}(A(x^{(k)}) - zI) \le \epsilon \right\}$.
8: $\quad$ **Return** $\widehat{x} \leftarrow x^{(k)}$, $\widehat{z} \leftarrow z^{(k)}$ if convergence occurred.
9: $\quad v^{(k)} \leftarrow$ a unit right singular vector corresponding to $\sigma_{\min}(A(x^{(k)}) - z^{(k)} I)$.
10: $\quad V_k \leftarrow \mathrm{orth}\left([V_{k-1} \;\; v^{(k)}] \right)$ and $\mathcal{V}_k \leftarrow \mathrm{Col}(V_k)$.
11: **end for**

reduced problem is solved in line 6, in particular, a global minimizer $\widetilde{x}$ is found for a reduced problem. Then a rightmost point $\widetilde{z}$ of $\Lambda_\epsilon(\widetilde{x})$ is determined in line 7. Finally, the subspace is expanded with the inclusion of a right singular vector corresponding to $\sigma_{\min}(A(\widetilde{x}) - \widetilde{z}I)$ in line 10. Note that the optimization problems in lines 2, 6 and 7 are nonconvex, and can have more than one global optimizers; argmin and argmax in these lines refer to any global minimizer and any global maximizer.

Determining the rightmost point of $\Lambda_\epsilon(\widetilde{x})$ involves the large-scale matrix-valued function $A(x)$, and is usually the most expensive step computationally in a subspace iteration. For this task, we usually benefit in practice from the approach in [27], an approach that combines the globally-convergent criss-cross algorithm [17] for computing the pseudospectral abscissa with a subspace framework. On the other hand, the small-scale reduced pseudspectral abscissa minimization problem is usually cheap to solve. For this task, we employ "eigopt", the globally convergent algorithm in [33], if there is only one parameter, or otherwise, if the matrix-valued function depends on multiple parameters, we use "GRANSO" [18]. The latter approach "GRANSO" is based on an adaptation of BFGS with a proper line search for nonsmooth optimization

problems, so can converge to local minimizers rather than global minimizers, yet better suited for problems with several parameters. On the other hand, "eigopt" is based on piecewise quadratic approximations of the objective, and guaranteed to converge globally provided a lower bound on the second derivatives is provided accurately as a parameter, but prohibitively expensive when there are more than a few parameters. Last but not the least, let us also note that the right singular vector of the large-scale matrix in line 9 should normally be calculated iteratively, for instance by means of "ARPACK" [28]. Further implementation details of the subspace framework, including the condition to check convergence used in practice in line 8, are described in Section 6.1.

2.1. Basic Results Regarding the Subspace Framework. Next, we present two basic results concerning the subspace framework that will be crucial in the convergence analysis.

The first result is the monotonicity property with respect to the subspace $\mathcal{V}$. We refer to Lemma 3.1 and succeeding arguments in [27] for a proof.

LEMMA 2.1 (Monotonicity). *Let $\mathcal{V}, \mathcal{W}$ be two subspaces of $\mathbb{C}^n$ such that $\mathcal{V} \subseteq \mathcal{W}$, and V, W be matrices whose columns form orthonormal bases for $\mathcal{V}, \mathcal{W}$. Then the following assertions hold:*

(i) $\sigma_{\min}(A(x)-zI) \leq \sigma_{\min}(A^W(x)-zW) \leq \sigma_{\min}(A^V(x)-zV) \ \ \forall z \in \mathbb{C}, \forall x \in \Omega.$

(ii) $\Lambda_\epsilon^{\mathcal{V}}(x) \subseteq \Lambda_\epsilon^{\mathcal{W}}(x) \subseteq \Lambda_\epsilon(x) \quad \forall x \in \Omega.$

(iii) $\alpha_\epsilon^{\mathcal{V}}(x) \leq \alpha_\epsilon^{\mathcal{W}}(x) \leq \alpha_\epsilon(x) \quad \forall x \in \Omega.$

The next result concerns the interpolation properties between the full and reduced problems, and is borrowed from [27, Lemma 3.2], where a proof is also provided.

LEMMA 2.2. *For a given $\widetilde{x} \in \Omega$, and given subspace $\mathcal{V}$, the following are equivalent:*

(i) $\alpha_\epsilon^{\mathcal{V}}(\widetilde{x}) = \alpha_\epsilon(\widetilde{x})$.

(ii) *The subspace $\mathcal{V}$ contains a right singular vector corresponding to $\sigma_{\min}(A(\widetilde{x}) - \widetilde{z}I)$ for some $\widetilde{z} \in \Lambda_\epsilon(\widetilde{x})$ with $\mathrm{Re}(\widetilde{z}) = \alpha_\epsilon(\widetilde{x})$.*

Global convergence of the proposed subspace framework as well as its rapid convergence can be attributed to the interpolation properties between the full and the reduced problems. In a subsequent subsection, we establish interpolation properties between $\alpha_\epsilon(\cdot)$ and $\alpha_\epsilon^{\mathcal{V}_k}(\cdot)$ as well as between their first derivatives for the subspaces $\mathcal{V}_k$ generated by Algorithm 2.1. Before stating the result formally, we first focus on the formulas for the derivatives of $\alpha_\epsilon(\cdot)$ and $\alpha_\epsilon^{\mathcal{V}}(\cdot)$ for a given subspace $\mathcal{V}$ in the next subsection.

2.2. Derivatives of the Pseudospectral Abscissa. We view the pseudospectral abscissa as a constrained optimization problem. Formally, letting $A(x,z) := A(x)-(z_1+\mathrm{i}z_2)I$ and $\sigma(x,z) := \sigma_{\min}(A(x,z))$ for $z = (z_1, z_2) \in \mathbb{R}^2$, the ϵ-pseudospectral abscissa $\alpha_\epsilon(x)$ can be expressed as

$$\alpha_\epsilon(x) \ = \ \max\{z_1 \ \mid \ z = (z_1, z_2) \in \mathbb{R}^2, \ \ \sigma(x,z) - \epsilon \leq 0\}. \tag{2.4}$$

We consider the Langrangian function

$$\mathcal{L}(x, z, \mu) \ := \ z_1 - \mu(\sigma(x,z) - \epsilon) \tag{2.5}$$

associated with (2.4), where $\mu \geq 0$ is the Lagrange multiplier corresponding to the constraint $\sigma(x,z) - \epsilon \leq 0$. For a given x, we denote the optimal $z = (z_1, z_2)$

and the corresponding μ for the optimization problem in (2.4), assuming they are unique, with $z(x) = (z_1(x), z_2(x))$ and $\mu(x)$, respectively. Moreover, we make use of the notations $y = (z, \mu)$ and $y(x) = (z(x), \mu(x))$. In the subsequent discussions, $\sigma_{z_1}(\cdot), \sigma_{z_2}(\cdot), \mathcal{L}_{z_1}(\cdot), \mathcal{L}_{z_2}(\cdot)$ represent the partial derivatives of $\sigma(\cdot)$, $\mathcal{L}(\cdot)$ with respect to z_1, z_2. For $u, w \in \{x, y\}$, the notations $\nabla_u \mathcal{L}(\cdot)$ and $\nabla^2_{uw}\mathcal{L}(\cdot)$ correspond to the gradient of $\mathcal{L}(\cdot)$ with respect to u and Hessian of $\mathcal{L}(\cdot)$ for which the partial derivatives are first taken with respect to u then with respect to w.

DEFINITION 2.3. *We call $x \in \Omega$ a nondegenerate point if*
(i) *the optimal z of (2.4), denoted as $z(x)$, is unique,*
(ii) *the singular value $\sigma(x, z(x))$ of $A(x, z(x))$ is simple, and*
(iii) *$\nabla_{yy}\mathcal{L}(x, y(x))$ is nonsingular.*

For part **(iii)** of the definition above, we remark that its parts **(i)-(ii)** ensure the uniqueness of $\mu(x)$, and so the uniqueness of $y(x)$ (i.e., see the proof of Theorem 2.4 in Appendix A).

The next theorem below can be regarded as a generalization of the part of [11, Theorem 8.1] that concerns the differentiability and first derivative of the map $X \in \mathbb{C}^{n\times n} \mapsto \alpha_\epsilon(X)$. Its proof exploits the optimality conditions for the constrained optimization problem (2.4), as well the implicit function function theorem, and given in the appendix.

THEOREM 2.4 (Derivatives of the Pseudospectral Abscissa Function). *Let $\widetilde{x} \in \Omega$ be a nondegenerate point. Then the function $x \mapsto \alpha_\epsilon(x)$ is real-analytic at $x = \widetilde{x}$ with*

$$\nabla \alpha_\epsilon(\widetilde{x}) \;=\; \left[\begin{array}{ccc} \mathrm{Re}\left(\frac{u^*\left[\frac{\partial A}{\partial x_1}(\widetilde{x})\right]v}{u^*v}\right) & \cdots & \mathrm{Re}\left(\frac{u^*\left[\frac{\partial A}{\partial x_d}(\widetilde{x})\right]v}{u^*v}\right) \end{array} \right]^T \tag{2.6}$$

where u, v consist of a pair of consistent unit left, right singular vectors, respectively, corresponding to $\sigma(\widetilde{x}, z(\widetilde{x}))$. Moreover,

$$\nabla^2 \alpha_\epsilon(\widetilde{x}) = \nabla^2_{xx}\mathcal{L}(\widetilde{x}, y(\widetilde{x})) - \nabla^2_{xy}\mathcal{L}(\widetilde{x}, y(\widetilde{x})) \cdot [\nabla^2_{yy}\mathcal{L}(\widetilde{x}, y(\widetilde{x}))]^{-1}[\nabla^2_{xy}\mathcal{L}(\widetilde{x}, y(\widetilde{x}))]^T \tag{2.7}$$

where

$$\begin{aligned}
[\nabla^2_{xx}\mathcal{L}(\widetilde{x}, y(\widetilde{x}))]_{j\ell} &= -\mu(\widetilde{x}) \left\{ \frac{\partial^2}{\partial x_j \partial x_\ell}[\sigma(x, z)]\Big|_{\widetilde{x}, z(\widetilde{x})} \right\} \qquad j, \ell = 1, 2, \cdots, d, \\
[\nabla^2_{xy}\mathcal{L}(\widetilde{x}, y(\widetilde{x}))]_{j\ell} &= -\mu(\widetilde{x}) \left\{ \frac{\partial^2}{\partial x_j \partial y_\ell}[\sigma(x, z)]\Big|_{\widetilde{x}, z(\widetilde{x})} \right\} \qquad j = 1, 2, \cdots d, \;\; \ell = 1, 2, 3, \\
[\nabla^2_{yy}\mathcal{L}(\widetilde{x}, y(\widetilde{x}))]_{j\ell} &= -\frac{\partial^2}{\partial y_j \partial y_\ell}[\mu \cdot \sigma(x, z)]\Big|_{\widetilde{x}, z(\widetilde{x}), \mu(\widetilde{x})} \qquad j, \ell = 1, 2, 3,
\end{aligned}$$

$(y_1, y_2, y_3) = (z_1, z_2, \mu)$, and $\mu(\widetilde{x}) = -1/(u^*v)$.

If we repeat the arguments above for the reduced pseudospectral abscissa function $\alpha^{\mathcal{V}}_\epsilon(\cdot)$ for a given subspace $\mathcal{V}$, analogous formulas for the derivatives of $\alpha^{\mathcal{V}}_\epsilon(\cdot)$ can be obtained in terms of a matrix V whose columns form an orthonormal basis for $\mathcal{V}$. Now $\alpha^{\mathcal{V}}_\epsilon(\cdot)$ can be expressed as the optimization problem

$$\alpha^{\mathcal{V}}_\epsilon(x) \;=\; \max\{z_1 \;\mid\; z = (z_1, z_2) \in \mathbb{R}^2, \;\; \sigma^{\mathcal{V}}(x, z) - \epsilon \le 0\}, \tag{2.8}$$

and the Lagrangian for the optimization problem associated with $\alpha^{\mathcal{V}}_\epsilon(\cdot)$ takes the form

$$\mathcal{L}^{\mathcal{V}}(x, z, \mu) \;:=\; z_1 - \mu(\sigma^{\mathcal{V}}(x, z) - \epsilon), \tag{2.9}$$

where $\sigma^{\mathcal{V}}(x,z) := \sigma_{\min}(A^V(x,z))$ and $A^V(x,z) := A^V(x) - zV$. We denote the optimal z and the corresponding Lagrange multiplier μ for the optimization problem in (2.8), assuming their uniqueness, by $z^{\mathcal{V}}(x)$ and $\mu^{\mathcal{V}}(x)$, and also let $y^{\mathcal{V}}(x) = (z^{\mathcal{V}}(x), \mu^{\mathcal{V}}(x))$ at a given $x \in \Omega$.

DEFINITION 2.5. *We call $x \in \Omega$ a nondegenerate point of the restriction of $A(\cdot)$ to a subspace $\mathcal{V}$ if*

- **(i)** *the optimal z of (2.8), denoted as $z^{\mathcal{V}}(x)$, is unique,*
- **(ii)** *the singular value $\sigma^{\mathcal{V}}(x, z^{\mathcal{V}}(x))$ of $A^V(x, z^{\mathcal{V}}(x))$ is simple for any matrix V whose columns form an orthonormal basis for $\mathcal{V}$, and*
- **(iii)** *$\nabla_{yy}\mathcal{L}^{\mathcal{V}}(x, y^{\mathcal{V}}(x))$ is nonsingular.*

Applying steps analogous to the ones in the proof of Theorem 2.4 yield the following result.

THEOREM 2.6 (Derivatives of the Reduced Pseudospectral Abscissa Function). *Let $\mathcal{V}$ be a subspace of $\mathbb{C}^n$, and $\widetilde{x} \in \Omega$ be a nondegenerate point of the restriction of $A(\cdot)$ to $\mathcal{V}$. Then the function $x \mapsto \alpha_\epsilon^{\mathcal{V}}(x)$ is real-analytic at $x = \widetilde{x}$ with*

$$\nabla \alpha_\epsilon^{\mathcal{V}}(\widetilde{x}) \;=\; \left[\; \mathrm{Re}\left(\frac{(u^{\mathcal{V}})^* \left[\frac{\partial A^V}{\partial x_1}(\widetilde{x})\right] v^{\mathcal{V}}}{(u^{\mathcal{V}})^* V v^{\mathcal{V}}}\right) \quad \cdots \quad \mathrm{Re}\left(\frac{(u^{\mathcal{V}})^* \left[\frac{\partial A^V}{\partial x_d}(\widetilde{x})\right] v^{\mathcal{V}}}{(u^{\mathcal{V}})^* V v^{\mathcal{V}}}\right) \;\right]^T \tag{2.10}$$

where V is a matrix whose columns form an orthonormal basis for $\mathcal{V}$, and $u^{\mathcal{V}}, v^{\mathcal{V}}$ form a pair of consistent unit left, unit right singular vectors corresponding to $\sigma^{\mathcal{V}}(\widetilde{x}, z^{\mathcal{V}}(\widetilde{x}))$. Furthermore,

$$\begin{aligned} \nabla^2 \alpha_\epsilon^{\mathcal{V}}(\widetilde{x}) \;=\; & \nabla^2_{xx}\mathcal{L}^{\mathcal{V}}(\widetilde{x}, y^{\mathcal{V}}(\widetilde{x})) \;- \\ & \nabla^2_{xy}\mathcal{L}^{\mathcal{V}}(\widetilde{x}, y^{\mathcal{V}}(\widetilde{x}))[\nabla^2_{yy}\mathcal{L}^{\mathcal{V}}(\widetilde{x}, y^{\mathcal{V}}(\widetilde{x}))]^{-1}[\nabla^2_{xy}\mathcal{L}^{\mathcal{V}}(\widetilde{x}, y^{\mathcal{V}}(\widetilde{x}))]^T, \end{aligned} \tag{2.11}$$

where

$$[\nabla^2_{xx}\mathcal{L}^{\mathcal{V}}(\widetilde{x}, y^{\mathcal{V}}(\widetilde{x}))]_{j\ell} \;=\; -\mu^{\mathcal{V}}(\widetilde{x}) \left\{ \frac{\partial^2}{\partial x_j \partial x_\ell}[\sigma^{\mathcal{V}}(x,z)]\Big|_{\widetilde{x}, z^{\mathcal{V}}(\widetilde{x})} \right\} \quad j,\ell = 1,2,\cdots,d,$$

$$[\nabla^2_{xy}\mathcal{L}^{\mathcal{V}}(\widetilde{x}, y^{\mathcal{V}}(\widetilde{x}))]_{j\ell} \;=\; -\mu^{\mathcal{V}}(\widetilde{x}) \left\{ \frac{\partial^2}{\partial x_j \partial y_\ell}[\sigma^{\mathcal{V}}(x,z)]\Big|_{\widetilde{x}, z^{\mathcal{V}}(\widetilde{x})} \right\} \quad j = 1,2,\cdots d, \;\; \ell = 1,2,3,$$

$$[\nabla^2_{yy}\mathcal{L}^{\mathcal{V}}(\widetilde{x}, y^{\mathcal{V}}(\widetilde{x}))]_{j\ell} \;=\; -\frac{\partial^2}{\partial y_j \partial y_\ell}[\mu \cdot \sigma^{\mathcal{V}}(x,z)]\Big|_{\widetilde{x}, z^{\mathcal{V}}(\widetilde{x}), \mu^{\mathcal{V}}(\widetilde{x})} \qquad j,\ell = 1,2,3,$$

and $\mu^{\mathcal{V}}(\widetilde{x}) = -1/\big((u^{\mathcal{V}})^* V v^{\mathcal{V}}\big)$.

Remark. Expressions for the third and higher-order derivatives of $x \mapsto \alpha_\epsilon(x)$ at a nondegenerate point $\widetilde{x} \in \Omega$ can be derived by differentiating (A.3) further. Similarly, the higher-order derivatives of $x \mapsto \alpha_\epsilon^{\mathcal{V}}(x)$ at a nondegenerate point of the restriction of $A(\cdot)$ to $\mathcal{V}$ can be obtained by differentiating the Lagrangian in (2.9) repeatedly.

2.3. Hermite Interpolation of the Pseudospectral Abscissa. The next result concerns the Hermite interpolation properties between the full pseudospectral abscissa function $\alpha_\epsilon(\cdot)$ and its reduced counterpart $\alpha_\epsilon^{\mathcal{V}_k}(\cdot)$ for a subspace $\mathcal{V}_k$ generated by Algorithm 2.1.

LEMMA 2.7 (Hermite Interpolation). *The following assertions hold regarding the subspace $\mathcal{V}_k$ for every integer $k \geq 2$ and the points $x^{(\ell)}$, $z^{(\ell)}$ for $\ell = 2, \ldots, k$ generated by Algorithm 2.1:*

(i) $\alpha_\epsilon(x^{(\ell)}) = \alpha_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$.
(ii) *The point* $z^{(\ell)}$ *is a rightmost point of* $\Lambda_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$.
(iii) *If* $x^{(\ell)}$ *is a nondegenerate point, then* $\alpha_\epsilon(x)$ *and* $\alpha_\epsilon^{\mathcal{V}_k}(x)$ *are differentiable at* $x = x^{(\ell)}$, *and* $\nabla\alpha_\epsilon(x^{(\ell)}) = \nabla\alpha_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$.

Proof. **(i)** From lines 9, 10 of Algorithm 2.1, we have $v^{(\ell)} \in \mathcal{V}_k$, where $v^{(\ell)}$ is a right singular vector corresponding to $\sigma_{\min}(A(x^{(\ell)}) - z^{(\ell)}I)$ for the rightmost point $z^{(\ell)}$ in $\Lambda_\epsilon(x^{(\ell)})$, for $\ell = 2, \dots k$. As a result, the assertion follows from Lemma 2.2.

(ii) As $v^{(\ell)} \in \mathcal{V}_k$, there exists a unit vector a such that $v^{(\ell)} = V_k a$, and so

$$\begin{aligned}\epsilon \geq \sigma_{\min}(A(x^{(\ell)}) - z^{(\ell)}I) &= \|(A(x^{(\ell)}) - z^{(\ell)})V_k a\|_2 \\ &\geq \sigma_{\min}(A(x^{(\ell)})V_k - z^{(\ell)}V_k).\end{aligned}$$

This shows that $z^{(\ell)} \in \Lambda_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$, and so $\alpha_\epsilon(x^{(\ell)}) = \mathrm{Re}(z^{(\ell)}) \leq \alpha_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$. But the reverse inequality $\alpha_\epsilon(x^{(\ell)}) \geq \alpha_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$ also holds due to Lemma 2.1 implying $\mathrm{Re}(z^{(\ell)}) = \alpha_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$, i.e., $z^{(\ell)}$ is a rightmost point in $\Lambda_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$.

(iii) Differentiability of $\alpha_\epsilon(x)$ at $x = x^{(\ell)}$ is immediate, i.e., $x^{(\ell)}$ is nondegenerate, so, from Definition 2.3, the rightmost point $z^{(\ell)}$ in $\Lambda_\epsilon(x^{(\ell)})$ is unique, and the smallest singular value of $A(x^{(\ell)}) - z^{(\ell)}I$ is simple.

Let us show that $\alpha_\epsilon^{\mathcal{V}_k}(x)$ is also differentiable at $x = x^{(\ell)}$. From part (ii), $z^{(\ell)}$ is a rightmost point in $\Lambda_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$. We claim that $z^{(\ell)}$ is the unique rightmost point in $\Lambda_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$. Suppose otherwise for the sake of contradiction, that is there exists $\widetilde{z} \neq z^{(\ell)}$ such that $\widetilde{z} \in \Lambda_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$ and $\alpha_\epsilon^{\mathcal{V}_k}(x^{(\ell)}) = \mathrm{Re}(z^{(\ell)}) = \mathrm{Re}(\widetilde{z})$. But then, by Lemma 2.1, we have $\epsilon \geq \sigma_{\min}(A^{V_k}(x^{(\ell)}) - \widetilde{z}V_k) \geq \sigma_{\min}(A(x^{(\ell)}) - \widetilde{z}I)$, so $\widetilde{z} \in \Lambda_\epsilon(A(x^\ell))$. Consequently, $\widetilde{z}$ and $z^{(\ell)}$ are rightmost points in $\Lambda_\epsilon(x^{(\ell)})$, contradicting the assumption that $x^{(\ell)}$ is nondegenerate.

Furthermore, if $\sigma_{\min}(A^{V_k}(x^{(\ell)}) - z^{(\ell)}V_k)$ is not simple, then there exist two unit orthonormal vectors $a, \widetilde{a}$ that satisfy

$$\begin{aligned}\sigma_{\min}(A(x^{(\ell)}) - z^{(\ell)}I) &= \sigma_{\min}(A^{V_k}(x^{(\ell)}) - z^{(\ell)}V_k) \\ &= \|(A^{V_k}(x^{(\ell)}) - z^{(\ell)}V_k)a\|_2 = \|(A^{V_k}(x^{(\ell)}) - z^{(\ell)}V_k)\widetilde{a}\|_2.\end{aligned}$$

But then $\omega = V_k a$ and $\widetilde{\omega} = V_k\widetilde{a}$ are orthonormal vectors satisfying

$$\sigma_{\min}(A(x^{(\ell)}) - z^{(\ell)}I) = \|(A(x^{(\ell)}) - z^{(\ell)})\omega\|_2 = \|(A(x^{(\ell)}) - z^{(\ell)})\widetilde{\omega}\|_2,$$

which contradicts with the simplicity of $\sigma_{\min}(A(x^{(\ell)}) - z^{(\ell)}I)$.

The arguments above show that $z^{(\ell)}$ is the unique rightmost point in $\Lambda_\epsilon^{\mathcal{V}_k}(x^{(\ell)})$, and that $\sigma_{\min}(A^{V_k}(x^{(\ell)}) - z^{(\ell)}V_k)$ is simple. Consequently, $\alpha_\epsilon^{\mathcal{V}_k}(x)$ is differentiable at $x = x^{(\ell)}$.

Finally, we show that the gradients of $\alpha_\epsilon(x)$ and $\alpha_\epsilon^{\mathcal{V}_k}(x)$ are equal at $x = x^{(\ell)}$. Let u, v be a pair of unit left and unit right singular vectors corresponding to $\sigma := \sigma_{\min}(A(x^{(\ell)}) - z^{(\ell)}I)$ satisfying

$$(A(x^{(\ell)}) - z^{(\ell)}I)v = \sigma u \qquad \text{and} \qquad u^*(A(x^{(\ell)}) - z^{(\ell)}I) = \sigma v^*.$$

Since $v \in \mathcal{V}_k$, there exists a unit vector a such that $v = V_k a$. Then the equations above can be rewritten in terms of a as follows:

$$(A^{V_k}(x^{(\ell)}) - z^{(\ell)}V_k)a = \sigma u \qquad \text{and} \qquad u^*(A^{V_k}(x^{(\ell)}) - z^{(\ell)}V_k) = \sigma a^*.$$

This means that u, a are a pair of unit left and unit right singular vectors corresponding to $\sigma_{\min}(A^{V_k}(x^{(\ell)}) - z^{(\ell)}V_k)$.

Using the analytical formulas derived in the previous subsection for the derivatives of $\alpha_\epsilon(\cdot)$ and $\alpha_\epsilon^{\mathcal{V}_k}()$, specifically using (2.6) and (2.10), we obtain

$$\frac{\partial \alpha_\epsilon}{\partial x_j}(x^{(\ell)}) = \mathrm{Re}\left(\frac{u^*\left[\frac{\partial A}{\partial x_j}(x^{(\ell)})\right]v}{u^*v}\right)$$
$$= \mathrm{Re}\left(\frac{u^*\left[\frac{\partial A^{V_k}}{\partial x_j}(x^{(\ell)})\right]a}{u^*V_k a}\right) \;=\; \frac{\partial \alpha_\epsilon^{\mathcal{V}_k}}{\partial x_j}(x^{(\ell)})$$

for $j = 1, \ldots, d$. ∎

3. Global Convergence. We now relate the sequence $x^{(2)}, x^{(3)}, x^{(4)}, \ldots$ generated by Algorithm 2.1 and the global minimizer of $\alpha_\epsilon(x)$ over all $x \in \underline{\Omega}$. The first result below asserts that if any two iterates are equal, then global convergence is achieved.

THEOREM 3.1. *If two points $x^{(\ell)}, x^{(k)}$ with $2 \le \ell < k$ generated by Algorithm 2.1 are equal, then $x^{(\ell)}$ is a global minimizer of $\alpha_\epsilon(x)$ over all $x \in \underline{\Omega}$.*

Proof. It follows from the monotonicity property (i.e., Lemma 2.1) that

$$\alpha_\epsilon^{\mathcal{V}_\ell}(x^{(k)}) \;\le\; \alpha_\epsilon^{\mathcal{V}_{k-1}}(x^{(k)}) \;=\; \min_{x\in\underline{\Omega}} \alpha_\epsilon^{\mathcal{V}_{k-1}}(x) \;\le\; \min_{x\in\underline{\Omega}} \alpha_\epsilon(x).$$

Moreover, an implication of the interpolation property (i.e., part **(i)** of Lemma 2.7) is that

$$\min_{x\in\underline{\Omega}} \alpha_\epsilon(x) \;\le\; \alpha_\epsilon(x^{(\ell)}) \;=\; \alpha_\epsilon^{\mathcal{V}_\ell}(x^{(\ell)}) \;=\; \alpha_\epsilon^{\mathcal{V}_\ell}(x^{(k)}),$$

where the last equality is due to $x^{(\ell)} = x^{(k)}$. Combining the inequalities above, $\min_{x\in\underline{\Omega}} \alpha_\epsilon(x) = \alpha_\epsilon^{\mathcal{V}_\ell}(x^{(k)}) = \alpha_\epsilon(x^{(\ell)})$ as desired. ∎

Remark. The subspace $\mathcal{V}_\ell$ is a subspace of $\mathbb{C}^n$ and contains $\mathcal{V}_{\ell-1}$ for every $\ell > 2$. As a result, we must have $\mathcal{V}_\ell = \mathcal{V}_{\ell-1}$ (so that $x^{(\ell+1)} = x^{(\ell)}$) for some $\ell > 2$. Theorem 3.1 implies $x^{(\ell)}$ for such an ℓ is a global minimizer of $\alpha_\epsilon(x)$ over $x \in \underline{\Omega}$.

3.1. Infinite Dimensional Setting. The global convergence of the subspace framework can remarkably be extended even to the infinite dimensional setting when the subspaces can grow arbitrarily as we show next. Here, we would like to remark that some of the robust stabilization problems originate from discretizations of infinite dimensional differential operators, e.g., discretizing the Orr-Sommerfeld equation that depends viscosity and density leads to finite dimensional autonomous systems whose robust stability is a concern. The analysis here indicates that the global convergence would be attained even if we were operating on an infinite discretization. Also, the map $z \mapsto A(x)z$ on $\mathbb{C}^n$ in finite dimension can be extended to infinite dimension by considering $(z, y) \mapsto (A(x)z, y)$ acting on a space of infinite sequences. Hence, the infinite-dimensional setting in this subsection encompasses the finite-dimensional setting as well.

Formally, following the practices in [23, 32], throughout the rest of this subsection $A_1, \ldots A_\kappa$ in (1.2), rather than representing $n \times n$ matrices, represent bounded linear operators on $\ell_2(\mathbb{N})$, which denotes the Hilbert space of square summable complex

infinite sequences equipped with the inner product $\langle v, w\rangle = \sum_{j=1}^{\infty} \overline{v}_j w_j$ and the norm $\|v\|_2 = \sqrt{\sum_{j=1}^{\infty} |v_j|^2}$. The operator-valued function $x \mapsto A(x)$ defined as in (1.2) for given real-analytic functions $f_1, \ldots, f_\kappa$ replaces the matrix-valued function in the finite dimensional setting. We assume $A(x) - zI$ has a singular value decomposition of the form

$$A(x) - zI \;=\; \sum_{j=1}^{\infty} \sigma_j \langle w_j, \cdot\rangle u_j$$

for every $x \in \Omega$ and $z \in \mathbb{C}$, where $\sigma_j \leq \sigma_{j+1}$ for all $j \geq 1$. The reduced matrix-valued function is now replaced by the parameter dependent linear map $z \in \mathbb{C}^r \mapsto A^V(x)z \in \ell_2(\mathbb{N})$, where $A^V(x)$ is still of the form (2.1) and $z \mapsto Vz$ is defined by $Vz := \sum_{j=1}^{r} z_j v_j$ for an orthonormal basis $\{v_1, \ldots, v_r\}$ of the subspace $\mathcal{V}$ of $\ell_2(\mathbb{N})$. Since $A(x)V - zV$ has finite rank, it has a singular value decomposition. The smallest singular values of $A(x,z) = A(x) - zI$ and $A^V(x,z) = A(x)V - zV$ are real analytic at $(\widetilde{x}, \widetilde{z})$ provided they are simple at $(\widetilde{x}, \widetilde{z})$ [24, p. 370, Theorem 1.8]. The definitions of $\Lambda_\epsilon(x), \Lambda_\epsilon^{\mathcal{V}}(x)$ and $\alpha_\epsilon(x), \alpha_\epsilon^{\mathcal{V}}(x)$ remain intact other than the change that $A(x)$ and $A^V(x)$ in these definitions are now operators and linear maps instead of matrices.

Algorithm 2.1 remains the same by again noting that parameter-dependent matrices are now parameter-dependent operators or linear maps. All of the basic results in Section 2.1, 2.2 and 2.3, including the monotonicity and Hermite interpolations results, hold after replacing the instances of the inner product $x^*y = \sum_{j=1}^{n} \overline{x}_j y_j$ on $\mathbb{C}^n$ for $x, y \in \mathbb{C}^n$ with the inner product $\langle x, y\rangle = \sum_{j=1}^{\infty} \overline{x}_j y_j$ on $\ell_2(\mathbb{N})$ for $x, y \in \ell_2(\mathbb{N})$. In particular, the gradient of the pseudospectral abscissa at a nondegenerate point $\widetilde{x}$ is

$$\nabla \alpha_\epsilon(\widetilde{x}) \;=\; \left[\; \mathrm{Re}\left(\frac{\left\langle u, \left[\frac{\partial A}{\partial x_1}(\widetilde{x})\right] v\right\rangle}{\langle u, v\rangle}\right) \quad \cdots \quad \mathrm{Re}\left(\frac{\left\langle u, \left[\frac{\partial A}{\partial x_d}(\widetilde{x})\right] v\right\rangle}{\langle u, v\rangle}\right) \;\right]^T$$

for a consistent pair u, v of unit left, right singular vectors corresponding to the smallest singular value of $A(\widetilde{x}) - z(\widetilde{x})I$. On the other hand, the gradient of the reduced pseudospectral abscissa at a nondegenerate point $\widetilde{x}$ for the restriction of $A(\cdot)$ to $\mathcal{V}$ is

$$\nabla \alpha_\epsilon^{\mathcal{V}}(\widetilde{x}) \;=\; \left[\; \mathrm{Re}\left(\frac{\left\langle u^{\mathcal{V}}, \left[\frac{\partial A^V}{\partial x_1}(\widetilde{x})\right] v^{\mathcal{V}}\right\rangle}{\langle u^{\mathcal{V}}, V v^{\mathcal{V}}\rangle}\right) \quad \cdots \quad \mathrm{Re}\left(\frac{\left\langle u^{\mathcal{V}}, \left[\frac{\partial A^V}{\partial x_d}(\widetilde{x})\right] v^{\mathcal{V}}\right\rangle}{\langle u^{\mathcal{V}}, V v^{\mathcal{V}}\rangle}\right) \;\right]^T$$

for a consistent pair $u^{\mathcal{V}}, v^{\mathcal{V}}$ of unit left, unit right singular vectors corresponding to the smallest singular value of $A^V(\widetilde{x}) - z^{\mathcal{V}}(\widetilde{x})V$.

As in the proof of Lemma 2.7, if $u^{(k)}$ and $v^{(k)}$ denote a consistent pair of unit left and right singular vectors corresponding to $\sigma_{\min}(A(x^{(k)}) - z^{(k)}I)$, then $u^{(k)}$ and $a^{(k)}$ satisfying $v_k = V_k a^{(k)}$ form a consistent pair of left and right singular vectors corresponding to $\sigma_{\min}(A^{V_k}(x^{(k)}) - z^{(k)}V_k)$. As argued in Section 2.2, assuming $\sigma_{\min}(A(x^{(k)}) - z^{(k)}I)$ is simple, by exploiting the fact that $z^{(k)}$ is the rightmost point in $\Lambda_\epsilon(x^{(k)})$, and employing the associated first order optimality conditions, $\left\langle u^{(k)}, v^{(k)}\right\rangle$ is real and nonzero, indeed $\left\langle u^{(k)}, v^{(k)}\right\rangle < 0$. In terms of the reduced problem, this amounts to

$$0 > \left\langle u^{(k)}, v^{(k)}\right\rangle = \left\langle u^{(k)}, V_k a^{(k)}\right\rangle = -\frac{\partial}{\partial z_1}\left[\sigma_{\min}\left(A^{V_k}(x, z)\right)\right]\Big|_{x^{(k)}, z^{(k)}}.$$

As argued in part (ii) of Lemma 2.7, $z^{(k)}$ above is also the rightmost point in $\Lambda_\epsilon^{\mathcal{V}_k}(x^{(k)})$.

Indeed, letting $u^{\mathcal{V}_k}(x)$, $a^{\mathcal{V}_k}(x)$ be unit consistent left, right singular vectors corresponding to $\sigma_{\min}(A^{V_k}(x) - z^{\mathcal{V}_k}(x)V_k)$ with $z^{\mathcal{V}_k}(x)$ representing the rightmost point in $\Lambda_\epsilon^{\mathcal{V}_k}(x)$, the property

$$0 \;>\; \left\langle u^{\mathcal{V}_k}(x), V_k a^{\mathcal{V}_k}(x)\right\rangle \tag{3.1}$$

holds uniformly for every $x \in \Omega$ that is a nondegenerate point for the restriction of $A(\cdot)$ to $\mathcal{V}_k$ as we argue next. Indeed, let $\widetilde{x} \in \Omega$ be such a nondegenerate point. By applying the first-order optimality conditions to the Lagrangian $\mathcal{L}^{\mathcal{V}_k}(x, z, \mu) := z_1 - \mu(\sigma^{\mathcal{V}_k}(x, z) - \epsilon)$, we deduce

$$\begin{aligned} 0 \;&=\; 1 - \mu^{\mathcal{V}_k}(\widetilde{x}) \operatorname{Re}\left(\left\langle u^{\mathcal{V}_k}(\widetilde{x}), \frac{\partial}{\partial z_1}\left[A^{V_k}(x,z)\right]\Big|_{\widetilde{x}, z^{\mathcal{V}_k}(\widetilde{x})} a^{\mathcal{V}_k}(\widetilde{x})\right\rangle\right) \\ &=\; 1 + \mu^{\mathcal{V}_k}(\widetilde{x}) \operatorname{Re}(\left\langle u^{\mathcal{V}_k}(\widetilde{x}), V_k a^{\mathcal{V}_k}(\overline{x})\right\rangle) \end{aligned}$$

and

$$\begin{aligned} 0 \;&=\; -\mu^{\mathcal{V}_k}(\widetilde{x}) \operatorname{Re}\left(\left\langle u^{\mathcal{V}_k}(\widetilde{x}), \frac{\partial}{\partial z_2}\left[A^{V_k}(x,z)\right]\Big|_{\widetilde{x}, z^{\mathcal{V}_k}(\widetilde{x})} a^{\mathcal{V}_k}(\widetilde{x})\right\rangle\right) \\ &=\; -\mu^{\mathcal{V}_k}(\widetilde{x}) \operatorname{Im}(\left\langle u^{\mathcal{V}_k}(\widetilde{x}), V_k a^{\mathcal{V}_k}(\widetilde{x})\right\rangle). \end{aligned}$$

From the first equation, we have $\mu^{\mathcal{V}_k}(\widetilde{x}) \neq 0$. As a result, from the second equation, $\operatorname{Im}(\left\langle u^{\mathcal{V}_k}(x), V_k a^{\mathcal{V}_k}(x)\right\rangle) = 0$, which shows $\left\langle u^{\mathcal{V}_k}(\widetilde{x}), V_k a^{\mathcal{V}_k}(\widetilde{x})\right\rangle$ is real. Additionally, $\left\langle u^{\mathcal{V}_k}(\widetilde{x}), V_k a^{\mathcal{V}_k}(\widetilde{x})\right\rangle \neq 0$ due to the first equation. Moreover, as $z^{\mathcal{V}_k}(\widetilde{x})$ is the rightmost point in $\Lambda_\epsilon^{\mathcal{V}_k}(\widetilde{x})$, we have

$$0 \;\leq\; \frac{\partial}{\partial z_1}\left[\sigma_{\min}\left(A^{V_k}(x,z)\right)\right]\Big|_{\widetilde{x}, z^{\mathcal{V}_k}(\widetilde{x})} \;=\; \left\langle -u^{\mathcal{V}_k}(\widetilde{x}), V_k a^{\mathcal{V}_k}(\widetilde{x})\right\rangle.$$

Combining this with $\left\langle u^{\mathcal{V}_k}(\widetilde{x}), V_k a^{\mathcal{V}_k}(\widetilde{x})\right\rangle \neq 0$ yields $0 > \left\langle u^{\mathcal{V}_k}(\widetilde{x}), V_k a^{\mathcal{V}_k}(\widetilde{x})\right\rangle$.

The uniform negativity feature in (3.1) is essential for the global convergence of the subspace framework. We indeed assume slightly more, namely a uniform negative local upper bound on the terms $\left\langle u^{\mathcal{V}_k}(x), V_k a^{\mathcal{V}_k}(x)\right\rangle$ near $x^{(k)}$ as stated formally below. Throughout the rest of this section and next section, $\mathcal{B}(c, r)$ for a given $c \in \mathbb{R}^q$ and $r \in \mathbb{R}$ denotes the closed ball

$$\mathcal{B}(c, r) \;:=\; \{\, \widetilde{c} \in \mathbb{R}^q \;\mid\; \|\widetilde{c} - c\|_2 \;\leq\; r \,\},$$

which reduces to the closed interval $[c - r, c + r]$ in case $q = 1$.

Assumption 3.2. *There exist $\varphi > 0$ and $\beta < 0$ such that for all $k \geq 2$ the following inequality holds:*

$$\beta \;>\; \sup_{x \,\in\, \Omega^{(k)} \cap \mathcal{B}(x^{(k)}, \varphi)} \left\langle u^{\mathcal{V}_k}(x), V_k a^{\mathcal{V}_k}(x)\right\rangle,$$

where $\Omega^{(k)}$ denotes the subset of Ω consisting of nondegenerate points for the restriction of $A(\cdot)$ to $\mathcal{V}_k$.

The following uniform local Lipschitz continuity result for $x \mapsto \alpha_\epsilon^{\mathcal{V}_k}(x)$ plays a prominent role when establishing the global convergence of the subspace framework.

LEMMA 3.3 (Uniform Local Lipschitz Continuity). *Suppose Assumption 3.2 is satisfied. There exist real scalars $\zeta > 0$ and $\varphi > 0$ such that for all $k \geq 2$ we have $\left\|\nabla \alpha_\epsilon^{\mathcal{V}_k}(x)\right\| \leq \zeta$ for every $x \in \Omega \cap \mathcal{B}(x^{(k)}, \varphi)$ that is a nondegenerate point for the restriction of $A(\cdot)$ to $\mathcal{V}_k$.*

Consequently, $\alpha_\epsilon^{\mathcal{V}_k}(x)$ is uniformly locally Lipschitz, that is there exist $\zeta > 0$ and $\varphi > 0$ such that for all $k \geq 2$ the following assertion holds:

$$\left|\alpha_\epsilon^{\mathcal{V}_k}(\widehat{x}) - \alpha_\epsilon^{\mathcal{V}_k}(\widetilde{x})\right| \leq \zeta \|\widehat{x} - \widetilde{x}\|_2 \quad \forall \widehat{x}, \widetilde{x} \in \Omega \cap \mathcal{B}(x^{(k)}, \varphi).$$

Proof. Let $\varphi > 0$ and $\beta < 0$ be as in Assumption 3.2, and $\widetilde{x} \in \Omega \cap \mathcal{B}(x^{(k)}, \varphi)$ be a nondegenerate point for the restriction of $A(\cdot)$ to $\mathcal{V}_k$. The partial derivatives of $x \mapsto \alpha_\epsilon^{\mathcal{V}_k}(x)$ at $x = \widetilde{x}$ are given by

$$\begin{aligned}\frac{\partial \alpha_\epsilon^{\mathcal{V}_k}}{\partial x_j}(\widetilde{x}) &= \operatorname{Re}\left(\frac{\left\langle u^{\mathcal{V}_k}(\widetilde{x}), \left[\frac{\partial A^{V_k}}{\partial x_j}(\widetilde{x})\right] a^{\mathcal{V}_k}(\widetilde{x})\right\rangle}{\langle u^{\mathcal{V}_k}(\widetilde{x}), V_k a^{\mathcal{V}_k}(\widetilde{x})\rangle}\right) \\ &= \operatorname{Re}\left(\sum_{\ell=1}^{\kappa} \frac{\partial f_\ell(\widetilde{x})}{\partial x_j} \frac{\langle u^{\mathcal{V}_k}(\widetilde{x}), A_\ell V_k a^{\mathcal{V}_k}(\widetilde{x})\rangle}{\langle u^{\mathcal{V}_k}(\widetilde{x}), V_k a^{\mathcal{V}_k}(\widetilde{x})\rangle}\right)\end{aligned}$$

for $j = 1, \ldots, d$, which yields

$$\begin{aligned}\left|\frac{\partial \alpha_\epsilon^{\mathcal{V}_k}}{\partial x_j}(\widetilde{x})\right| &\leq \sum_{\ell=1}^{\kappa} \left|\frac{\partial f_\ell(\widetilde{x})}{\partial x_j}\right| \left|\frac{\langle u^{\mathcal{V}_k}(\widetilde{x}), A_\ell V_k a^{\mathcal{V}_k}(\widetilde{x})\rangle}{\langle u^{\mathcal{V}_k}(\widetilde{x}), V_k a^{\mathcal{V}_k}(\widetilde{x})\rangle}\right| \\ &\leq \sum_{\ell=1}^{\kappa} \left|\frac{\partial f_\ell(\widetilde{x})}{\partial x_j}\right| \frac{\|A_\ell\|_2}{|\beta|},\end{aligned}$$

where the last summation is independent of $\mathcal{V}_k$. This shows the existence of a ζ such that for all $k \geq 2$ the bound $\left\|\nabla \alpha_\epsilon^{\mathcal{V}_k}(x)\right\| \leq \zeta$ holds for every $x \in \Omega \cap \mathcal{B}(x^{(k)}, \varphi)$ that is a nondegenerate point for the restriction of $A(\cdot)$ to $\mathcal{V}_k$.

Due to the boundedness of the gradients, the uniform Lipschitz continuity follows from the continuity of $\alpha_\epsilon^{\mathcal{V}_k}(x)$, and a simple application of the mean value theorem. ∎

Now we are ready for the global convergence result. Its proof follows the footsteps of that for [23, Theorem 3.1], yet we present it below for completeness.

THEOREM 3.4 (Global Convergence). *Suppose Assumption 3.2 is satisfied. Then the following hold for Algorithm 2.1 in the infinite dimensional setting:*

- ***(i)*** *The limit of every convergent subsequence of $\left\{x^{(k)}\right\}$ is a global minimizer of $\alpha_\epsilon(x)$ over all $x \in \underline{\Omega}$.*
- ***(ii)*** $\lim_{k\to\infty} \alpha_\epsilon^{\mathcal{V}_k}(x^{(k+1)}) = \lim_{k\to\infty} \min_{x\in\underline{\Omega}} \alpha_\epsilon^{\mathcal{V}_k}(x) = \min_{x\in\underline{\Omega}} \alpha_\epsilon(x).$

Proof. Let $\left\{x^{(\ell_k)}\right\}$ be a convergent subsequence of $\left\{x^{(k)}\right\}$. By the monotonicity property (i.e., extension of Lemma 2.1, part **(iii)** to the infinite dimensional setting), we have

$$\min_{x\in\underline{\Omega}} \alpha_\epsilon(x) \geq \min_{x\in\underline{\Omega}} \alpha_\epsilon^{\mathcal{V}_{\ell_{k+1}-1}}(x) = \alpha_\epsilon^{\mathcal{V}_{\ell_{k+1}-1}}(x^{(\ell_{k+1})}) \geq \alpha_\epsilon^{\mathcal{V}_{\ell_k}}(x^{(\ell_{k+1})}). \tag{3.2}$$

Additionally, due to interpolation property (i.e., extension of Lemma 2.7, part **(i)** to the infinite dimensional setting),

$$\min_{x\in\underline{\Omega}} \alpha_\epsilon(x) \leq \alpha_\epsilon(x^{(\ell_k)}) = \alpha_\epsilon^{\mathcal{V}_{\ell_k}}(x^{(\ell_k)}). \tag{3.3}$$

By combining (3.2) and (3.3), we obtain

$$\alpha_\epsilon^{\mathcal{V}_{\ell_k}}(x^{(\ell_{k+1})}) \;\le\; \min_{x\in\underline{\Omega}} \alpha_\epsilon(x) \;\le\; \alpha_\epsilon^{\mathcal{V}_{\ell_k}}(x^{(\ell_k)}).$$

It follows from the uniform local Lipschitz continuity of $\alpha_\epsilon^{\mathcal{V}_k}(x)$ with respect to k (i.e., Lemma 3.3), and noting $\lim_{k\to\infty} \|x^{(\ell_{k+1})} - x^{(\ell_k)}\|_2 = 0$ as $\{x^{(\ell_k)}\}$ is convergent, that

$$\lim_{k\to\infty} \left|\alpha_\epsilon^{\mathcal{V}_{\ell_k}}(x^{(\ell_{k+1})}) - \alpha_\epsilon^{\mathcal{V}_{\ell_k}}(x^{(\ell_k)})\right| \;\le\; \lim_{k\to\infty} \zeta \left\|x^{(\ell_{k+1})} - x^{(\ell_k)}\right\|_2 \;=\; 0,$$

where ζ is the uniform Lipschitz constant in Lemma 3.3. This together with the interpolation property (i.e., Lemma 2.7 in infinite dimension) imply that

$$\lim_{k\to\infty} \alpha_\epsilon(x^{(\ell_k)}) \;=\; \lim_{k\to\infty} \alpha_\epsilon^{\mathcal{V}_{\ell_k}}(x^{(\ell_k)}) \;=\; \lim_{k\to\infty} \alpha_\epsilon^{\mathcal{V}_{\ell_k}}(x^{(\ell_{k+1})}) \;=\; \min_{x\in\underline{\Omega}} \alpha_\epsilon(x). \tag{3.4}$$

Using the continuity of $\alpha_\epsilon(x)$ completes the proof of part **(i)**.

Letting $\alpha_* := \min_{x\in\underline{\Omega}} \alpha_\epsilon(x)$, for $p > k$, using monotonicity, we have

$$\begin{aligned} \alpha_* \;&\ge\; \min_{x\in\underline{\Omega}} \alpha_\epsilon^{\mathcal{V}_p}(x) \;=\; \alpha_\epsilon^{\mathcal{V}_p}(x^{(p+1)}) \\ &\ge\; \alpha_\epsilon^{\mathcal{V}_k}(x^{(p+1)}) \;\ge\; \alpha_\epsilon^{\mathcal{V}_k}(x^{(k+1)}). \end{aligned}$$

This shows that the sequence $\left\{\alpha_\epsilon^{\mathcal{V}_k}(x^{(k+1)})\right\}$ is monotonically increasing and bounded from above by α_*, so it is convergent. Consider the subsequence $\left\{\alpha_\epsilon^{\mathcal{V}_{\ell_{k+1}-1}}(x^{(\ell_{k+1})})\right\}$ of $\left\{\alpha_\epsilon^{\mathcal{V}_k}(x^{(k+1)})\right\}$, which satisfies

$$\alpha_\epsilon^{\mathcal{V}_{\ell_k}}(x^{(\ell_{k+1})}) \;\le\; \alpha_\epsilon^{\mathcal{V}_{\ell_{k+1}-1}}(x^{(\ell_{k+1})}) \;\le\; \alpha_*.$$

Since we have $\lim_{k\to\infty} \alpha_\epsilon^{\mathcal{V}_{\ell_k}}(x^{(\ell_{k+1})}) = \alpha_*$ from the first part, in particular from (3.4), $\left\{\alpha_\epsilon^{\mathcal{V}_{\ell_{k+1}-1}}(x^{(\ell_{k+1})})\right\}$ also converges to α_*. Finally, the limit of the convergent sequence $\left\{\alpha_\epsilon^{\mathcal{V}_k}(x^{(k+1)})\right\}$ must be the same as the limit α_* of its subsequence. ∎

4. Rate of Convergence. In this section, we prove that the rate of convergence of Algorithm 2.1 when $d = 1$ is superlinear under mild assumptions. It seems possible to generalize the arguments and the result when $d > 1$ provided that additional singular vectors are put into the subspace at every iteration at points close to the interpolation point employed by Algorithm 2.1; this extension is similar to the extension of Algorithm 1 to Algorithm 2 in [23] in the context of minimizing the jth largest eigenvalue for a prescribed j. We first focus on the case $d = 1$ for the sake of simplicity. In a subsection at the end of the section, we formally introduce its extension to attain superlinear convergence in theory when $d > 1$, and outline why the convergence results for Algorithm 2.1 applies to this extension

Throughout, it is assumed that three consecutive iterates $x^{(k-1)}, x^{(k)}, x^{(k+1)}$ generated by Algorithm 2.1 for $d = 1$ are sufficiently close to x_*, a global minimizer of $\alpha_\epsilon(x)$ over $x \in \Omega$. An additional assumption that is kept throughout is that x_* is strictly in the interior of $\underline{\Omega}$. Moreover, by z_* we denote a point in $\Lambda(x_*)$ satisfying $\alpha_\epsilon(x_*) = \mathrm{Re}(z_*)$. Furthermore, we assume the nondegeneracy of the point x_*, which is stated formally below.

Assumption 4.1 (Nondegeneracy of Optimizer). *The point x_* is in the interior of $\underline{\Omega}$, and is a nondegenerate point, that is the following conditions are satisfied by x_*, z_*:*

(i) z_* *is the unique point in* $\Lambda_\epsilon(x_*)$ *such that* $\mathrm{Re}(z_*) = \alpha_\epsilon(x_*)$.
(ii) *the smallest singular value* $\sigma(x_*, z_*)$ *of* $A(x_*, z_*)$ *is simple.*
(iii) $\nabla_{yy}\mathcal{L}(x_*, y_*)$ *is invertible, where* $\mathcal{L}(x, y)$ *is the Lagrangian as in (2.5) with* $y = (z, \mu)$, *and* $y_* = (z_*, \mu_*)$ *with* $\mu_* \in \mathbb{R}$ *represents the optimal point satisfying* $\nabla_y \mathcal{L}(x_*, y_*) = 0$.

By Theorem 2.4, nondegeneracy assumption above for x_* guarantees that the pseudospectral abscissa map $x \mapsto \alpha_\epsilon(x)$ is real analytic at $x = x_*$. We keep another assumption throughout this section, which concerns the second derivatives of the Lagrangian and its reduced counterpart. Note that, throughout this section, $y_* = (z_*, \mu_*)$ with $\mu_* \in \mathbb{R}$ is as in Assumption 4.1, that is it represents the unique y_* satisfying $\nabla_y \mathcal{L}(x_*, y_*) = 0$.

Assumption 4.2 (Robust Nondegeneracy). *For a given constant* $\gamma > 0$, *the following assertions hold:*

$$\sigma_{\min}\left(\nabla^2_{yy}\mathcal{L}(x_*, y_*)\right) \geq \gamma \quad \text{and} \quad \sigma_{\min}\left(\nabla^2_{yy}\mathcal{L}^{\mathcal{V}_k}(x_*, y_*)\right) \geq \gamma.$$

We start our derivation of the rate of convergence with a lemma that asserts that the pseudospectral abscissa functions of the full and the reduced problems attained at a unique smooth optimizer (i.e., a unique right-most point in pseudospectra, that is smooth with respect to x) around x_* under Assumptions 4.1 and 4.2. We omit its proof, as the proof of its part **(i)** is similar to [23, Proposition 2.9], and proofs of parts **(ii)**-**(iii)** to [32, Lemma 15].

Lemma 4.3. *Suppose that Assumption 4.1 and 4.2 hold. There exist* $\nu_x, \nu_z, \nu_\mu > 0$ *independent of* k *such that* $\mathcal{B}(x_*, \nu_x) \subseteq \underline{\Omega}$ *that satisfy the following:*

(i) *The singular value functions* $\sigma(x, z)$ *and* $\sigma^{\mathcal{V}_k}(x, z)$ *are simple, and their first three derivatives in absolute value are bounded above by constants uniformly for all* $x \in \mathcal{B}(x_*, \nu_x)$, $z \in \mathcal{B}(z_*, \nu_z)$, *where the constants are independent of* k.
(ii) *There exists a real-analytic function*

$$x \in \mathcal{B}(x_*, \nu_x) \mapsto y(x) = (z(x), \mu(x)) \in \mathcal{B}(z_*, \nu_z) \times \mathcal{B}(\mu_*, \nu_\mu)$$

such that $\alpha_\epsilon(x) = \mathrm{Re}(z(x))$, *as well as*

$$\nabla_y \mathcal{L}(x, y(x)) = 0 \quad \text{and} \quad \sigma_{\min}(\nabla^2_{yy}\mathcal{L}(x, y(x))) \geq \gamma/2 \tag{4.1}$$

for all $x \in \mathcal{B}(x_*, \nu_x)$.
(iii) *There exists a real-analytic function*

$$x \in \mathcal{B}(x_*, \nu_x) \mapsto y^{\mathcal{V}_k}(x) = (z^{\mathcal{V}_k}(x), \mu^{\mathcal{V}_k}(x)) \in \mathcal{B}(z_*, \nu_z) \times \mathcal{B}(\mu_*, \nu_\mu)$$

such that $\alpha^{\mathcal{V}_k}_\epsilon(x) = \mathrm{Re}(z^{\mathcal{V}_k}(x))$, *and*

$$\nabla_y \mathcal{L}^{\mathcal{V}_k}(x, y^{\mathcal{V}_k}(x)) = 0 \quad \text{and} \quad \sigma_{\min}(\nabla^2_{yy}\mathcal{L}^{\mathcal{V}_k}(x, y^{\mathcal{V}_k}(x))) \geq \gamma/2 \tag{4.2}$$

for all $x \in \mathcal{B}(x_*, \nu_x)$.

For the main rate-of-convergence result, as stated formally below, we also assume that the angle between the left and right singular vectors of $\sigma^{\mathcal{V}_k}(x, z^{\mathcal{V}_k}(x))$ obeys a certain bound on $\mathcal{B}(x_*, \nu_x)$, the interval in Lemma 4.3. This is similar to Assumption 3.2 in the global convergence proof. To this end, letting $u(x)$ and $v(x)$ be a pair of consistent unit left and right singular vectors corresponding to $\sigma(x, z(x))$, following the

arguments in Section 2.2, in particular by an application of the first order optimality conditions to the constrained optimization characterization in (2.4) of $\alpha_\epsilon(x)$, we have $u(x)^*v(x) < 0$ for all $x \in \mathcal{B}(x_*, \nu_x)$ so that

$$\underline{\beta} := \max_{x \in \mathcal{B}(x_*,\nu_x)} u(x)^*v(x) \;<\; 0. \tag{4.3}$$

Similarly, the counterparts of the arguments in Section 3.1 leading to (3.1) but now in finite dimension yield $u^{\mathcal{V}_k}(x)^*V_k v^{\mathcal{V}_k}(x) < 0$ for all $x \in \mathcal{B}(x_*, \nu_x)$, where $u^{\mathcal{V}_k}(x)$ and $v^{\mathcal{V}_k}(x)$ denote a consistent pair of unit left and right singular vectors corresponding to $\sigma^{\mathcal{V}_k}(x, z^{\mathcal{V}_k}(x))$. Below, we require slightly more, i.e., there is a negative upper bound independent of k on $u^{\mathcal{V}_k}(x)^*V_k v^{\mathcal{V}_k}(x)$ over all $x \in \mathcal{B}(x_*, \nu_x)$.

Assumption 4.4. *For a given constant β small enough in absolute value and such that $\underline{\beta} \le \beta < 0$, the subspace $\mathcal{V}_k$ is such that*

$$\beta \;\ge\; \max_{x \in \mathcal{B}(x_*,\nu_x)} u^{\mathcal{V}_k}(x)^*V_k v^{\mathcal{V}_k}(x),$$

where $\mathcal{B}(x_*, \nu_x)$ is as in Lemma 4.3.

The next result concerns the uniform boundedness of the derivatives of $\alpha_\epsilon(x)$ and $\alpha_\epsilon^{\mathcal{V}_k}(x)$.

Lemma 4.5. *Suppose that the conditions of Assumptions 4.1, 4.2 and 4.4 hold. There exists ν_x independent of k such that $\mathcal{B}(x_*, \nu_x) \subseteq \underline{\Omega}$, and the following assertions hold:*

- **(i)** *The pseudospectral functions $\alpha_\epsilon(x)$, $\alpha_\epsilon^{\mathcal{V}_k}(x)$ are real analytic at all $x \in \mathcal{B}(x_*, \nu_x)$.*
- **(ii)** *The first three derivatives of $\alpha_\epsilon(x)$ and $\alpha_\epsilon^{\mathcal{V}_k}(x)$ in absolute value are bounded above by constants uniformly for all $x \in \mathcal{B}(x_*, \nu_x)$, where the constants are independent of k.*

Proof. The real-analyticity of $\alpha_\epsilon(x)$ and $\alpha_\epsilon^{\mathcal{V}_k}(x)$ uniformly in an interval $\mathcal{B}(x_*, \nu_x)$ for some ν_x that is independent of k is immediate from Lemma 4.3.

The proof of boundedness of the first derivative of $\alpha_\epsilon(x)$ is similar to that in Lemma 3.3 for $\alpha_\epsilon^{\mathcal{V}_k}(x)$. In particular, noting $u(x)^*v(x) \le \beta < 0$ for $x \in \mathcal{B}(x_*, \nu_x)$ due to (4.3), and using the formulas (2.6) for the derivatives of $x \mapsto \alpha_\epsilon(x)$, at any $\widetilde{x} \in \mathcal{B}(x_*, \nu_x)$, we have

$$\alpha_\epsilon'(\widetilde{x}) \;=\; \mathrm{Re}\left(\sum_{\ell=1}^{\kappa} \frac{\mathrm{d}f_\ell(\widetilde{x})}{\mathrm{d}x} \frac{u(\widetilde{x})^*A_\ell v(\widetilde{x})}{u(\widetilde{x})^*v(\widetilde{x})}\right) \;\;\Longrightarrow\;\; |\alpha_\epsilon'(\widetilde{x})| \;\le\; \sum_{\ell=1}^{\kappa} \left|\frac{\mathrm{d}f_\ell(\widetilde{x})}{\mathrm{d}x}\right| \frac{\|A_\ell\|_2}{|\beta|},$$

implying the boundedness of $|\alpha_\epsilon'(x)|$ on $\mathcal{B}(x_*, \nu_x)$, as f_ℓ is real analytic for $\ell = 1, \ldots, \kappa$. As for the second derivatives of $\alpha_\epsilon(x)$, first observe

$$\mu(x) = -\frac{1}{u(x)^*v(x)} \;\le\; \frac{1}{-\beta} \tag{4.4}$$

for all $x \in \mathcal{B}(x_*, \nu_x)$. Moreover, from the continuity of the singular values and Assumption 4.2, we have

$$\sigma_{\min}(\nabla_{yy}^2 \mathcal{L}(x, y(x))) \;\ge\; \gamma/2 \tag{4.5}$$

for all $x \in \mathcal{B}(x_*, \nu_x)$, where γ is as in Assumption 4.2, by choosing ν_x smaller if necessary. From formula (2.7) for the second derivative of $\alpha_\epsilon(x)$ but in finite dimension, we have

$$|\alpha_\epsilon''(x)| \;\le\; |\mathcal{L}_{xx}(x, y(x))| + \left\|\nabla_{xy}^2 \mathcal{L}(x, y(x))\right\|_2^2 \|[\nabla_{yy}^2 \mathcal{L}(x, y(x))]^{-1}\|_2, \tag{4.6}$$

where $\mathcal{L}_{xx}(x,y)$ is the second derivative of $\mathcal{L}(x,y)$ with respect to x. Now, since the derivatives $\mathcal{L}_{xx}(x,y(x))$ and $\nabla^2_{xy}\mathcal{L}(x,y(x))$ can be expressed fully in terms of $\mu(x)$ and the partial derivatives of $\sigma(x,y(x))$, the result follows from (4.4), (4.5) and part **(i)** of Lemma 4.3.

The uniform boundedness of the third derivatives of $\alpha_\epsilon(x)$, and the boundedness of the derivatives of the reduced pseduospectral abscissa function can be shown in a similar way. ∎

Interpolation properties between $\alpha_\epsilon(x)$ and $\alpha_\epsilon^{\mathcal{V}_k}(x)$ and their first derivatives hold at $x^{(k)}$. Even if these interpolation properties do not extend to the second derivatives, the second derivatives must be close at $x^{(k)}$ as shown next.

LEMMA 4.6 (Proximity of the Second Derivatives). *Suppose that Assumptions 4.1, 4.2 and 4.4 are satisfied. There exists a constant $C > 0$ independent of k such that*

$$\left|\alpha_\epsilon''(x^{(k)}) - [\alpha_\epsilon^{\mathcal{V}_k}]''(x^{(k)})\right| \leq C\left|x^{(k)} - x^{(k-1)}\right|. \tag{4.7}$$

Proof. We assume without loss of generality that $x^{(k-1)}$, $x^{(k)}$ are strictly inside $\mathcal{B}(x_*,\nu_x)$ (i.e., the interval in Lemma 4.5), where $\alpha_\epsilon(x)$ and $\alpha_\epsilon^{\mathcal{V}_k}(x)$ are real analytic with uniform bounds on their first three derivatives independent of k.

Setting $h := x^{(k-1)} - x^{(k)}$, let us introduce the functions

$$l(t) := \alpha_\epsilon(x^{(k)} + th) \text{ and } l_k(t) := \alpha_\epsilon^{\mathcal{V}_k}(x^{(k)} + th) \tag{4.8}$$

for $t \in [0,1]$. By applying the Taylor's theorem with third order remainder to $l(t)$ and $l_k(t)$ on the interval $(0,1)$, we obtain

$$\begin{aligned} l(1) &= l(0) + l'(0) + l''(0)/2 + l'''(\eta)/6, \\ l_k(1) &= l_k(0) + l_k'(0) + l_k''(0)/2 + l_k'''(\eta_k)/6 \end{aligned}$$

for some constants $\eta, \eta_k \in (0,1)$. Notice that $l(1) = l_k(1)$, $l(0) = l_k(0)$, and $l'(0) = l_k'(0)$ due to Lemma 2.7. Consequently,

$$\begin{aligned} \frac{\alpha_\epsilon''(x^{(k)})h^2 - [\alpha^{\mathcal{V}_k}]''(x^{(k)})h^2}{2} &= \frac{l''(0) - l_k''(0)}{2} \\ &= \frac{l_k'''(\eta_k) - l'''(\eta)}{6} \\ &= \frac{[\alpha^{\mathcal{V}_k}]'''(x^{(k)} + \eta_k h)h^3 - \alpha_\epsilon'''(x^{(k)} + \eta h)h^3}{6}. \end{aligned} \tag{4.9}$$

As $x^{(k)} + \eta_k h, x^{(k)} + \eta h \in \mathcal{B}(x_*,\nu_x)$, the third derivatives $[\alpha^{\mathcal{V}_k}]'''(x^{(k)} + \eta_k h)$ and $\alpha_\epsilon'''(x^{(k)} + \eta h)$ on the righthand side of (4.9) in absolute value are bounded from above by a uniform constant U. Hence, we deduce from (4.9) that

$$\left|\alpha_\epsilon''(x^{(k)}) - [\alpha_\epsilon^{\mathcal{V}_k}]''(x^{(k)})\right| \leq \frac{2U}{3}h = \frac{2U}{3}\left|x^{(k)} - x^{(k-1)}\right|$$

as desired. ∎

Now we are ready to state and prove the main result of this section, that is the superlinear convergence result regarding the iterates of Algorithm 2.1.

THEOREM 4.7 (Superlinear Convergence). *Suppose that Assumptions 4.1, 4.2 and 4.4 hold. Additionally, assume that* $\alpha_\epsilon''(x_*) \neq 0$. *Then, there exists a constant* $\Upsilon > 0$ *independent of* k *such that*

$$\frac{\left|x^{(k+1)} - x_*\right|}{\left|x^{(k)} - x_*\right| \max\{\left|x^{(k)} - x_*\right|, \left|x^{(k-1)} - x_*\right|\}} \leq \Upsilon. \tag{4.10}$$

Proof. Let ν_x be as in Lemma 4.5 so that $\alpha_\epsilon(x)$ and $\alpha_\epsilon^{\mathcal{V}_k}(x)$ are differentiable, indeed $\alpha_\epsilon''(x)$ and $[\alpha_\epsilon^{\mathcal{V}_k}]''(x)$ are Lipschitz continuous, inside $\mathcal{B}(x_*, \nu_x)$. Moreover, without loss of generality, assume $x^{(k-1)}, x^{(k)}, x^{(k+1)} \in \mathcal{B}(x_*, \nu_x)$.

By assumption $\alpha_\epsilon''(x_*) \neq 0$, so, by employing Lemma 4.6, we can assume $x^{(k-1)}$, $x^{(k)}$ are close enough so that $[\alpha_\epsilon^{\mathcal{V}_k}]''(x_*) \neq 0$. Indeed, if necessary by choosing ν_x even smaller, we can assume $\alpha_\epsilon''(x) \neq 0$, $[\alpha_\epsilon^{\mathcal{V}_k}]''(x) \neq 0$ for all $x \in \mathcal{B}(x_*, \nu_x)$.

The superlinear convergence assertion in (4.10) now follows by expanding $\alpha_\epsilon'(x_*)$ about $x^{(k)}$ by employing Taylor's theorem with integral remainder. The derivation here is identical to that in Part 2 in the proof of [1, Theorem 3.3] by replacing $\sigma(\omega), \sigma_r(\omega)$ in that work with $\alpha_\epsilon(x), \alpha_\epsilon^{\mathcal{V}_k}(x)$, and $\omega_*, \omega_{r+1}, \omega_r, \omega_{r-1}$ in that work with $x_*, x^{(k+1)}, x^{(k)}, x^{(k-1)}$. ∎

In the context of a related subspace framework for maximizing the smallest eigenvalue of a matrix-valued function dependent on one parameter, the order of convergence is shown to be at least $1+\sqrt{2}$ in [26]. It is possible that the order of convergence of Algorithm 2.1 is also better than quadratic, which is suggested by the numerical experiments we have performed.

4.1. Extended Subspace Framework. When $d > 1$ the arguments above leading to the superlinear convergence assertion for Algorithm 2.1 fails, because the generalization of (4.7) that concerns the proximity of the second partial derivatives of $\alpha_\epsilon(x)$, $\alpha_\epsilon^{\mathcal{V}_k}(x)$ at $x^{(k)}$ does not have to hold.

A remedy to this issue is proposed in Algorithm 4.1. The difference of Algorithm 4.1 compared to Algorithm 2.1 is that additional right singular vectors at points close to $x^{(k)}$ are also included when forming the subspace $\mathcal{V}_k$. Let $h^{(k)} = \|x^{(k)} - x^{(k-1)}\|$, and $e_{pq} = 1/\sqrt{2}(e_p + e_q)$ if $p \neq q$ and $e_{pp} = e_p$, where e_p denotes the pth column of the $d \times d$ identity matrix. To be precise, in lines 11-17 of Algorithm 4.1, a right singular vector $v_{pq}^{(k)}$ corresponding to $\sigma_{\min}(A(x_{pq}^{(k)}) - z_{pq}^{(k)} I)$ is computed at $x_{pq}^{(k)} = x^{(k)} + h^{(k)} e_{pq}$ and $z_{pq}^{(k)}$, a rightmost point in $\Lambda_\epsilon(x_{pq}^{(k)})$, for $p = 1, \ldots d$, $q = p, \ldots, d$. In addition to $v^{(k)}$, all of these singular vectors are also included in the subspace $\mathcal{V}_k$. Clearly, this approach is practical when there are only a few parameters.

All of the results and derivations when deducing the superlinear convergence for Algorithm 2.1 when $d = 1$ hold for Algorithm 4.1 when $d > 1$ with minor modifications. To be specific, the Hermite interpolation property, that is Lemma 2.7 holds, not only at $x^{(\ell)}$ but also at nearby $x_{pq}^{(\ell)}$ for $p = 1, \ldots, d$, $q = p, \ldots, d$. Moreover, Lemma 4.3 extends without modification except that $\mathcal{B}(x_*, \nu_x)$ is now not an interval but a ball in $\mathbb{R}^d$, and Lemma 4.5 holds after replacing all instances of the derivatives with partial derivatives. In the proof of Lemma 4.6, we now assume without loss of generality that not only $x^{(k)}, x^{(k-1)}$ but also $x_{pq}^{(k)}$ for $p = 1, \ldots, d$, are contained in the ball $\mathcal{B}(x_*, \nu_x)$, and the proximity result

$$\left| \frac{\partial^2 \alpha_\epsilon}{\partial x_p \partial x_q}(x^{(k)}) - \frac{\partial^2 \alpha_\epsilon^{\mathcal{V}_k}}{\partial x_p \partial x_q}(x^{(k)}) \right| \leq C\|x^{(k)} - x^{(k-1)}\|$$

for a constant C independent of k is obtained by applying Taylor's theorem to $l_{pq}(t) := \alpha_\epsilon(x^{(k)} + th^{(k)}e_{pq})$ and $l_{k,pq}(t) := \alpha_\epsilon^{\mathcal{V}_k}(x^{(k)} + th^{(k)}e_{pq})$ for $p = 1, \ldots, d$, $q = p, \ldots, d$. Finally, Lemma 4.7, that is the main superlinear convergence result, extends for Algorithm 4.1 when $d > 1$ after replacing the instances of the absolute value in (4.10) with the norm, and the assumption $\alpha_\epsilon''(x_*) \neq 0$ with the invertibility of the Hessian $\nabla^2\alpha_\epsilon(x_*)$. The derivation is based on Taylor's theorem with integral remainder, and similar to the proof of Theorem 3.3 in [23].

Theorem 3.1 that concerns global convergence holds also for Algorithm 4.1, and Theorem 3.4 holds for the infinite-dimensional variant of Algorithm 4.1, as we still attain the monotonicity $\alpha_\epsilon^{\mathcal{V}_{k-1}}(x) \leq \alpha_\epsilon^{\mathcal{V}_k}(x)$ for all $x \in \Omega$, and Hermite interpolation at points $x^{(k)}$ with Algorithm 4.1.

Algorithm 4.1 The extended subspace framework to minimize $\alpha_\epsilon(x)$ over $\underline{\Omega}$

Input: The matrix-valued function $A(x)$, the feasible region $\underline{\Omega}$, and $\epsilon > 0$.
Output: An estimate $\widehat{x}$ for $\arg\min_{x\in\underline{\Omega}} \alpha_\epsilon(x)$, and $\widehat{z} \in \mathbb{C}$ that is an estimate for a globally rightmost point in $\Lambda_\epsilon(\widehat{x})$

$x_1^{(1)}, \ldots x_\eta^{(1)} \leftarrow$ initially chosen points in $\underline{\Omega}$.
$z_j^{(1)} \leftarrow \arg\max \left\{ \mathrm{Re}(z) \mid \sigma_{\min}(A(x_j^{(1)}) - zI) \leq \epsilon \right\}$ for $j = 1, \ldots, \eta$.
$v_j^{(1)} \leftarrow$ a right singular vector corr. to $\sigma_{\min}(A(x_j^{(1)}) - z_j^{(1)}I)$ for $j = 1, \ldots, \eta$.
$\mathcal{V}_1 \leftarrow \mathrm{span}\left\{ v_1^{(1)}, \ldots v_\eta^{(1)} \right\}$ and $V_1 \leftarrow$ an orthonormal basis for $\mathcal{V}_1$.
for $k = 2, 3, \ldots$ **do**
  $x^{(k)} \leftarrow \arg\min_{x\in\underline{\Omega}} \alpha_\epsilon^{\mathcal{V}_{k-1}}(x)$.
  $z^{(k)} \leftarrow \arg\max \left\{ \mathrm{Re}(z) \mid \sigma_{\min}(A(x^{(k)}) - zI) \leq \epsilon \right\}$.
  Return $\widehat{x} \leftarrow x^{(k)}$, $\widehat{z} \leftarrow z^{(k)}$ if convergence occurred.
  $v^{(k)} \leftarrow$ a unit right singular vector corresponding to $\sigma_{\min}(A(x^{(k)}) - z^{(k)}I)$.
  $h^{(k)} \leftarrow \|x^{(k)} - x^{(k-1)}\|$
  for $p = 1, \ldots, d$ **do**
    for $q = p, \ldots, d$ **do**
      $x_{pq}^{(k)} \leftarrow x^{(k)} + h^{(k)}e_{pq}$
      $z_{pq}^{(k)} \leftarrow \arg\max \left\{ \mathrm{Re}(z) \mid \sigma_{\min}(A(x_{pq}^{(k)}) - zI) \leq \epsilon \right\}$.
      $v_{pq}^{(k)} \leftarrow$ a unit right singular vector corresponding to $\sigma_{\min}(A(x_{pq}^{(k)}) - z_{pq}^{(k)}I)$.
    end for
  end for
  $V_k \leftarrow \mathrm{orth}\left([V_{k-1} \;\; v^{(k)} \;\; v_{11}^{(k)} \;\; \ldots v_{dd}^{(k)}] \right)$ and $\mathcal{V}_k \leftarrow \mathrm{Col}(V_k)$.
end for

5. Extensions to Real Pseudospectral Abscissa. The real ϵ-pseudospectrum, and real ϵ-pseudospectral abscissa of $A \in \mathbb{C}^{n\times n}$ for a given $\epsilon > 0$ are defined by

$$\Lambda_\epsilon^{\mathbb{R}}(A) := \{ z \in \mathbb{C} \mid z \in \Lambda(A + \Delta) \;\; \exists \Delta \in \mathbb{R}^{n\times n} \;\; \text{s.t.} \;\; \|\Delta\|_2 \leq \epsilon \},$$

and $\alpha_\epsilon^{\mathbb{R}}(A) := \max\left\{ \mathrm{Re}(z) \mid z \in \Lambda_\epsilon^{\mathbb{R}}(A) \right\}$. The ϵ-level-set characterizations [8]

$$\Lambda_\epsilon^{\mathbb{R}}(A) \;=\; \{ z \in \mathbb{C} \,|\, \mu(z) \leq \epsilon \}, \quad \alpha_\epsilon^{\mathbb{R}}(A) \;=\; \max\{ \mathrm{Re}(z) \in \mathbb{C} \,|\, \mu(z) \leq \epsilon \},$$

where

$$\begin{aligned} &\mu(z) := \max_{\gamma\in(0,1]} g(z,\gamma)\,, \\ &g(z,\gamma) := \sigma_{-2}\left(G(z,\gamma) := \begin{bmatrix} A-\mathrm{Re}(z)I & -\gamma\,\mathrm{Im}(z)I \\ \gamma^{-1}\,\mathrm{Im}(z)I & A-\mathrm{Re}(z)I \end{bmatrix}\right), \end{aligned} \tag{5.1}$$

and $\sigma_{-2}(\cdot)$ denotes the second smallest singular value of its matrix argument are useful for computational purposes. Fixed-point iterations based on low-rank dynamics are proposed in the literature for the computation of $\alpha_\epsilon^{\mathbb{R}}(A)$ [19, 36]. Moreover, a criss-cross type algorithm, and a subspace method for the computation of $\alpha_\epsilon^{\mathbb{R}}(A)$ are introduced in [31].

For a matrix-valued function $A(x)$ defined as in (1.2) and a prescribed $\epsilon > 0$, it may be of interest to minimize $\alpha_\epsilon^{\mathbb{R}}(A(x))$ over x in a compact set $\underline{\Omega}$ when the uncertainties are constrained to be real. A direct extension of Algorithm 2.1 would minimize a reduced counterpart of $\alpha_\epsilon^{\mathbb{R}}(A(x))$ such as the one in [31, Section 6], and locate the rightmost point z_* in $\Lambda_\epsilon^{\mathbb{R}}(A(x_*))$ at the minimizer x_* of the reduced problem. Then the subspace defining the reduced problem can be expanded with the inclusion of directions defined in terms of the singular vectors corresponding to the smallest two singular values of $G(z_*,\gamma_*)$ as in (5.1) with $A = A(x_*)$, where γ_* is the maximizer of $g(z_*,\gamma)$ over γ. The monotonicity properties analogous to the ones in Lemma 2.1 are immediate as argued in [31, Section 6]. However, it appears that the interpolation properties are lost with such a subspace framework; it seems that the interpolation properties do not have to hold even between the function μ in (5.1) and its reduced counterpart, as μ is the maximum of a singular value function rather than merely a singular value function. Extension of Algorithm 2.1 for the large-scale minimization of the real pseudospectral abscissa, or at least analyzing such an extension, is not straightforward, and goes beyond the scope of this work.

6. Numerical Results. We have implemented the proposed subspace framework, that is Algorithm 2.1, to minimize $\alpha_\epsilon(x)$ in Matlab. In this section, we perform numerical experiments using this implementation in Matlab 2020b on an an iMac with Mac OS 12.1 operating system, Intel® Core™ i5-9600K CPU and 32GB RAM. Our implementation of the framework seems to be converging rapidly even for multiparameter examples. Before focusing on numerical experiments on synthetic examples and benchmark examples taken from the *COMPl$_e$ib* collection [29] in Sections 6.2 and 6.3, we first summarize a few important implementation details in the next subsection.

6.1. Implementation Details.

6.1.1. Initial Interpolation Points. We choose the initial interpolation points $x_1,\ldots,x_\eta$ in line 1 of the subspace framework as follows unless they are made available explicitly. If there are multiple optimization parameters, then $x_1,\ldots,x_\eta$ are selected randomly in $\underline{\Omega}$. Otherwise, if there is only one optimization parameter, we choose them as equally-spaced points in $\underline{\Omega} := [L,U]$, i.e., $x_j := L+(j-1)\frac{U-L}{\eta-1}$ for $j = 1,\ldots,\eta$. In all of the experiments below, unless otherwise stated, the number of initial interpolation points is $\eta = 10$.

6.1.2. Termination Condition. In practice we terminate when the gap between the optimal values for the reduced optimization problems in two consecutive iterations is less than a prescribed tolerance. Formally, given a tolerance `tol`, we terminate at iteration $k \geq 3$ if

$$\alpha_\epsilon^{\mathcal{V}_{k-1}}(x^{(k)}) - \alpha_\epsilon^{\mathcal{V}_{k-2}}(x^{(k-1)}) \;<\; \texttt{tol}. \tag{6.1}$$

In practice we use $\texttt{tol} = 10^{-7}$. Additionally, we terminate if the number of subspace iterations exceeds a certain amount, which is 30 iterations in all of the numerical experiments below. But this second condition is never needed for the examples here as the condition in (6.1) with $\texttt{tol} = 10^{-7}$ is always met in fewer than 10 iterations.

6.1.3. Solution of the Reduced Optimization Problems. The minimization of $\alpha_\epsilon^{\mathcal{V}_{k-1}}(x)$ over $x \in \Omega$ in line 6 of Algorithm 2.1 is performed using either "eigopt" [33] or "GRANSO" [18].

If there is only one parameter, then we rely on "eigopt" as it converges globally provided a lower bound γ on the second derivatives of $\alpha_\epsilon^{\mathcal{V}_{k-1}}(x)$ where it is differentiable is chosen small enough. In all of the experiments below depending on one parameter, we set $\gamma = -400$, which seems to work well.

On the other hand, if there are multiple parameters, we depend on "GRANSO" for the solution of these reduced optimization problems. The reason is "eigopt" is usually slow when there are multiple optimization parameters, and the number of function evaluations needed to satisfy a certain accuracy increases quite rapidly with respect to the accuracy required. "GRANSO", on the other hand, can solve multiple-parameter problems quite efficiently. Yet, as it is based on BFGS and quasi-Newton methods, it can converge to a local minimizer that is not a global minimizer. For the numerical examples depending on multiple parameters that we report below, this local convergence issue does not cause any problem.

6.1.4. Objective for the Reduced Optimization Problems and Rectangular Pseudospectra. The objective for the reduced optimization problem at the kth subspace iteration is $\alpha_\epsilon^{\mathcal{V}_{k-1}}(x)$ meaning $\alpha_\epsilon^{\mathcal{V}_{k-1}}(x)$ needs to be computed at several x. This requires finding the rightmost point of the rectangular pseudospectrum

$$\Lambda_\epsilon(A^{V_{k-1}}(x)) \;=\; \{z \in \mathbb{C} \mid \sigma_{\min}(A^{V_{k-1}}(x) - zV_{k-1}) \leq \epsilon\}.$$

In general, finding the rightmost point of a rectangular psedospectrum of the form

$$\{z \in \mathbb{C} \mid \sigma_{\min}(E - zF) \leq \epsilon\}$$

for given matrices $E, F \in \mathbb{C}^{p\times q}$ with $p > q$ is a challenging problem. For instance, it does not seem easy to determine how many connected components such a rectangular pseudospectrum has, and where the connected components are in the complex plane.

To find the rightmost point of $\Lambda_\epsilon(A^{V_{k-1}}(x))$, we adopt the approach proposed in [27, Section 5.2], which is an extension of the quadratically convergent criss-cross algorithm for the ϵ-pseudospectral abscissa of a square matrix [17] to rectangular pencils. We remark that $A^{V_{k-1}}(x)$ and V_{k-1} are of size $n \times \widetilde{k}$, where $\widetilde{k} := k - 1 + \eta$. Hence, at first look they are not small scale. Yet, as suggested in [27, Section 5.2], a reduced QR factorization

$$\begin{bmatrix} V_{k-1} & A^{V_{k-1}}(x) \end{bmatrix} \;=\; Q \begin{bmatrix} \widetilde{B} & \widetilde{A} \end{bmatrix}$$

yields $\widetilde{A}, \widetilde{B} \in \mathbb{C}^{2\widetilde{k}\times\widetilde{k}}$ such that

$$\Lambda_\epsilon(A^{V_{k-1}}(x)) \;=\; \Lambda_\epsilon(\widetilde{A}, \widetilde{B}) := \{z \in \mathbb{C} \mid \sigma_{\min}(\widetilde{A} - z\widetilde{B}) \leq \epsilon\}.$$

To summarize, to compute $\alpha_\epsilon^{\mathcal{V}_{k-1}}(x)$ the rightmost point of $\Lambda_\epsilon(\widetilde{A}, \widetilde{B})$ depending on small matrices is found cheaply using a criss-cross algorithm as in [17].

One subtle issue is that it is essential for the criss-cross algorithm to locate the rightmost point in $\Lambda_\epsilon(\widetilde{A},\widetilde{B})$ to start iterating initially from a point in the rightmost connected component of $\Lambda_\epsilon(\widetilde{A},\widetilde{B})$. As a heuristic, it is proposed in [27, Section 5.2] to initialize the criss-cross algorithm with the rightmost eigenvalue λ of the pencil $L(s) = \widetilde{A}_1 - s\widetilde{B}_1$ satisfying $\sigma_{\min}(\widetilde{A} - \lambda\widetilde{B}) \le \epsilon$, where $\widetilde{A}_1, \widetilde{B}_1$ denote the upper $\widetilde{k} \times \widetilde{k}$ parts of $\widetilde{A}, \widetilde{B}$. The difficulty is that there may not be such an eigenvalue of $L(\cdot)$ satisfying $\sigma_{\min}(\widetilde{A} - \lambda\widetilde{B}) \le \epsilon$. As a safeguard, we additionally choose equally-spaced points $y_1, \ldots, y_s$ in a prescribed subinterval (for the examples below $[-2\mathrm{i}, 2\mathrm{i}]$) of the imaginary axis, then find the largest x such that $\sigma_{\min}(\widetilde{A}-(x+\mathrm{i}y_j)) = \epsilon$ for $j = 1, \ldots, s$. Such largest x corresponding to y_j, call it $x_{R,j}$, for $j = 1, \ldots, s$ is given by the imaginary part of one of the purely imaginary eigenvalues of the pencil

$$\mathcal{L}(s) \;=\; \begin{bmatrix} -y_j\widetilde{B}^* + \mathrm{i}\widetilde{A}^* & \epsilon I \\ -\epsilon I & y_j\widetilde{B} + \mathrm{i}\widetilde{A} \end{bmatrix} - s \begin{bmatrix} \widetilde{B}^* & 0 \\ 0 & \widetilde{B} \end{bmatrix}$$

with the largest imaginary part [27, Lemma 5.1]. To conclude, we initialize the criss-cross algorithm for rectangular pencils with the rightmost point among the points $x_{R,1}, \ldots, x_{R,s}$ and the point produced by the heuristic described above from [27, Section 5.2].

6.1.5. Finding the Rightmost Point of $\Lambda_\epsilon(x)$. The computation of the rightmost point of $\Lambda_\epsilon(x_j^{(1)})$ for $j = 1, 2, \ldots, \eta$, and $\Lambda_\epsilon(x^{(j)})$ for $j = 2, \ldots, k$ is required by the subspace framework to form the subspace $\mathcal{V}_k$. This is usually the most intensive computational task, as it involves finding the rightmost point in the ϵ-pseudospectral abscissa of the large matrix $A(x)$ at various x. For this purpose, we either employ the original criss-cross algorithm for the computation of the pseudospectral abscissa [17] if the size of $A(x)$ is less than or equal to a prescribed amount, or otherwise the subspace framework in [27] if the size of $A(x)$ is larger than the prescribed amount. In all of the experiments below, this prescribed size is chosen as 1000.

6.2. Synthetic Examples Depending on One Parameter. We first conduct numerical experiments with the synthetic examples used in [32, Section 7], that are publicly available[1]. These examples concern the distance to instability, which, for a matrix $M \in \mathbb{C}^{n\times n}$, is defined by

$$\mathcal{D}(M) \;:=\; \inf\{\, \|\Delta\|_2 \mid \Delta \in \mathbb{C}^{n\times n} \text{ s.t. } \Lambda(M+\Delta) \cap \mathbb{C}^+ \neq \emptyset \,\},$$

where $\mathbb{C}^+$ denotes the closed right-half of the complex plane. It follows from the definition of $\mathcal{D}(M)$ that, for every $\epsilon > 0$, we have

$$\mathcal{D}(M) \le \epsilon \quad\Longleftrightarrow\quad \alpha_\epsilon(M) \ge 0. \tag{6.2}$$

Each example in [32, Section 7] involves a matrix-valued function of the form $A(x) = A + xbc^T$ over $x \in \mathbb{R}$ in a prescribed interval for given $A \in \mathbb{R}^{n\times n}$, $b, c \in \mathbb{R}^n$. Specifically, the distance to instability for each example $A(x)$ is maximized over x in a prescribed interval $\mathcal{I}$, and the maximal value of the distance to instability, as well as the maximizer are reported. Letting x_* be the global maximizer of $\mathcal{D}(A(x))$ over $x \in \mathcal{I}$, it is immediate from (6.2) that $\min_{x\in\mathcal{I}} \alpha_\epsilon(A(x)) = \alpha_\epsilon(A(x_*)) = 0$ for $\epsilon = \mathcal{D}(A(x_*))$.

We illustrate the proposed subspace framework in Figure 6.1 to minimize $\alpha_\epsilon(x) := \alpha_\epsilon(A(x))$ for the example $A(x)$ in [32, Section 7] of order $n = 400$ over $x \in [-0.3, 0.2]$,

[1] http://home.ku.edu.tr/~emengi/software/max_di/Data_&_Updates.html

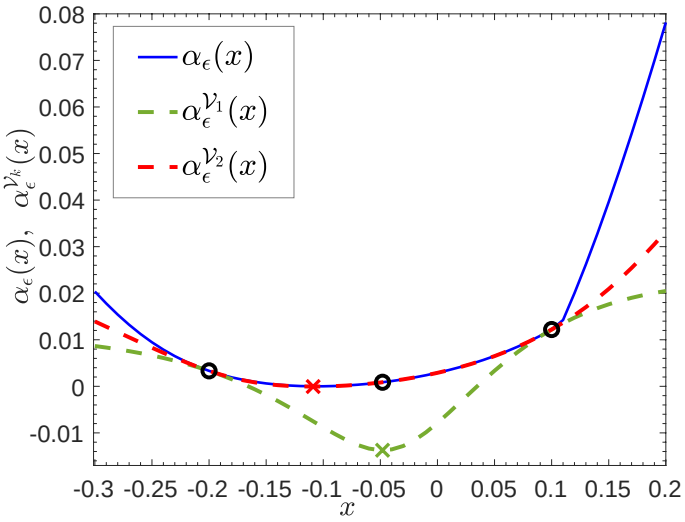

FIG. 6.1. *The progress of the subspace framework to minimize $\alpha_\epsilon(A(x))$ over $x \in [-0.3, 0.2]$ for the example $A(x)$ in [32, Section 7] of order $n = 400$, and with $\epsilon \approx \mathcal{D}(A(x_*))$. The black circles mark the interpolation points, whereas the crosses mark the global minimizers of $\alpha_\epsilon^{\mathcal{V}_1}(x)$ and $\alpha_\epsilon^{\mathcal{V}_2}(x)$.*

and with $\epsilon = 0.12870882 \approx \mathcal{D}(A(x_*))$. The initial subspace $\mathcal{V}_1$ is chosen as the 2-dimensional subspace so that Hermite interpolation is attained between $\alpha_\epsilon(x)$ and $\alpha_\epsilon^{\mathcal{V}_1}(x)$ at $x = -0.2, 0.1$. The reduced function $\alpha_\epsilon^{\mathcal{V}_1}(x)$ is minimized over $x \in [-0.3, 0.2]$; its minimizer turns out to be $x^{(2)} = -0.0480905$. Then the subspace $\mathcal{V}_1$ is expanded into $\mathcal{V}_2$ so that $\alpha_\epsilon^{\mathcal{V}_2}(x)$ interpolates $\alpha_\epsilon(x)$ at $x = x^{(2)}$. The global minimizer of $\alpha_\epsilon^{\mathcal{V}_2}(x)$ is already quite close to the actual minimizer of $\alpha_\epsilon(x)$ as can be seen in Figure 6.1. The subspace framework on this example terminates after 4 subspace iterations. The iterates of the subspace framework are given in Table 6.1. It appears from the second and third columns of this table that $x^{(k)}$ converges to the minimizer x_* of $\alpha_\epsilon(x)$ at a superlinear rate, consistent with the superlinear convergence assertion of Theorem 4.7. Also, as expected, the globally smallest value of $\alpha_\epsilon(A(x))$ is about 0, and the computed global minimizer $x_* = -0.1056316$ is about the same as the global maximizer of $\mathcal{D}(A(x))$ reported in [32, Section 7].

TABLE 6.1
The iterates of the subspace framework to minimize $\alpha_\epsilon(A(x))$ for the example $A(x)$ in [32, Section 7] of order $n = 400$, and with $\epsilon \approx \mathcal{D}(A(x_*))$.

k	$x^{(k+1)}$	$\lvert x^{(k+1)} - x_* \rvert$	$\alpha_\epsilon^{\mathcal{V}_k}(x^{(k+1)})$
1	-0.04808976223	0.05754188697	-0.01372736840
2	-0.10883718893	0.00320553973	-0.00005010004
3	-0.10563118290	0.00000004663	-0.00000001017
4	-0.10563164920	0.00000000000	-0.00000000753

Next we apply the subspace framework to minimize $\alpha_\epsilon(x)$ for the same example but for several values of ϵ ranging from 10^{-6} to 1. The computed minimal values and minimizers of $\alpha_\epsilon(x)$ are listed with respect to ϵ in Table 6.2. The optimal values appear to be in harmony with $\alpha_\epsilon(x) = 0$ for ϵ equal to the maximal distance to

TABLE 6.2

The application of the subspace framework to minimize the ϵ-pseudospectral abscissa of the example from [32, Section 7] of order $n = 400$ on the interval $[-0.3, 0.2]$ for various values of ϵ. The global minimizer x_*, corresponding minimal value of $\alpha_\epsilon(A(x))$, and number of subspace iterations are listed.

ϵ	x_*	$\alpha_\epsilon(x_*)$	iter	ϵ	x_*	$\alpha_\epsilon(x_*)$	iter
10^{-6}	-0.26107	-0.4381593	4	0.12	-0.09836	-0.0215016	2
10^{-5}	-0.26114	-0.4380593	4	0.2	-0.15639	0.1602082	2
10^{-4}	-0.26181	-0.4370651	4	0.4	-0.19641	0.5285811	2
10^{-3}	-0.26742	-0.4276521	4	0.6	-0.14273	0.8384027	2
10^{-2}	-0.27941	-0.3620685	3	0.8	-0.11798	1.1192159	2
10^{-1}	-0.07932	-0.0731799	2	1	-0.10290	1.3831928	2

instability attainable, which is about 0.12870882. For smaller values of ϵ, the optimal ϵ-pseudospectral abscissa is negative, whereas, for larger ϵ values, it is positive as expected. We especially remark that the subspace framework seems to work well for small values of ϵ such as 10^{-6}.

We have also performed experiments with the examples from [32, Section 7] of order $n = 200, 400, 800, 1200, 2000$ that concern the maximization of $\mathcal{D}(A(x))$ on the interval $[-3, 3]$, where $A(x) = A + xbc^T$ for given $A^{n\times n}, b, c \in \mathbb{R}^n$. In each case, we have minimized $\alpha_\epsilon(A(x))$ over $x \in [-3, 3]$ using the subspace framework for $\epsilon \approx \mathcal{D}(A(x_*))$ (to be precise for ϵ equal to the reported value of $\mathcal{D}(A(x_*))$ in [32, Section 7], which is eight decimal digit accurate). The results are listed in Table 6.3. The globally minimal value $\alpha_\epsilon(x_*)$ of $\alpha_\epsilon(x)$ is about 0 for each one of these examples as expected. Moreover, the global minimizers x_* of $\mathcal{D}(A(x))$ listed in the table are close to those reported in [32, Section 7]. The number of iterations until termination for each one of the examples is 3 to 6 indicating quick convergence. As for the runtimes, according to the table the main computational task that contributes to the runtime is the computation of the ϵ-pseudospectral abscissa of $A(x)$, which is required once per iteration as well as to form the initial subspaces. On the other hand, the reduced optimization problem that involves the minimization of $\alpha_\epsilon^{\mathcal{V}_k}(x)$ at the kth subspace iteration is relatively cheap to solve. These two computational tasks determine the total runtime.

The runtimes for a direct minimization (i.e., via `eigopt`) of $\alpha_\epsilon(x)$ without the subspace framework for minimization are given in the last column of Table 6.3. Here, $\alpha_\epsilon(x)$ is the objective, and needs to be computed for several values of x. Following the practice used in the subspace framework, $\alpha_\epsilon(x)$ is computed directly using the criss-cross algorithm [11] if the size of $A(x)$ is at most 1000 (i.e., in the first three cases in the table with $n = 200, 400, 800$). On the other hand, if the size of $A(x)$ is larger than 1000, then the subspace framework in [27] is employed to compute $\alpha_\epsilon(x)$. In all cases in the table, the direct minimization of $\alpha_\epsilon(x)$ takes considerably more time as compared to the minimization of $\alpha_\epsilon(x)$ via the proposed subspace framework.

6.3. Benchmark Examples. Several benchmark examples for the stabilization by static output feedback problem are provided in the *COMPl$_e$ib* collection [29]. The problems that we consider here taken from *COMPl$_e$ib* concern finding a $K \in \mathbb{R}^{m\times p}$ such that $A + BKC$ has all of its eigenvalues in the open left-half of the complex plane for given $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$. Minimizing the ϵ-pseudospectral abscissa of $A + BKC$ over K for a prescribed ϵ has also been suggested in [29] for the robust stabilization of the system.

We focus on three examples. The first one is about the stabilization of a single-

TABLE 6.3
The table concerns the application of the subspace framework to the examples from [32, Section 7] of order $n = 200, 400, 800, 1200, 2000$. In each case, we minimize $\alpha_\epsilon(A(x))$ over $x \in [-3, 3]$, where ϵ is the reported largest value of $\mathcal{D}(A(x))$ over $x \in [-3, 3]$. The last four columns list the total runtime (time), time for the reduced minimization problems (red), time to compute the rightmost point of $\Lambda_\epsilon(A(x))$ at various x (psa) for the subspace framework, and time to solve the minimization problem directly without using the subspace framework for minimization (direct) in seconds.

n	$\epsilon \approx \mathcal{D}(A(x_*))$	x_*	$\alpha_\epsilon(x_*)$	iter	time	red	psa	direct
200	0.01839422	0.09439	0	5	17.7	7	10.6	50.6
400	0.12870882	−0.10566	0	4	78.9	23.5	55.1	1547.1
800	0.11545563	−0.05851	0.0000010	4	241.3	24.7	214.9	5713.7
1200	0.07941192	−0.48625	0.0000003	3	181.4	64.1	116.2	8732.8
2000	0.08436380	−0.04766	0.0000005	6	737.9	24.5	708.5	7574.6

TABLE 6.4
The iterates of the subspace framework to minimize $\alpha_\epsilon(x)$ over $x \in [-1, 1]$ for the NN18 example.

k	$x^{(k+1)}$	$\alpha_\epsilon^{\mathcal{V}_k}(x^{(k+1)})$
1	0.23190	−1.0680310
2	−1.00000	−0.9149600
3	−1.00000	−0.9149600

input-single-output system so that K is a scalar meaning there is only one parameter to be optimized, whereas the remaining two examples depend on multiple optimization parameters. Especially, the first and third examples involve relatively large matrices so that a direct minimization of the ϵ-pseudospectral abscissa requires considerably more time, which we report below. In all of the examples, ϵ is chosen as 0.2.

NN18 *($n = 1006$, $m = p = 1$).* This example involves the stabilization of $A+xbc^T$ over $x \in \mathbb{R}$. We minimize $\alpha_\epsilon(x)$ for $\epsilon = 0.2$ over $x \in [-1, 1]$ by employing the subspace framework. The computed global minimizer is $x_* = -1$, and the computed globally smallest ϵ-pseudospectral abscissa is $\alpha_\epsilon(x_*) = -0.9149600$. The correctness of these computed values can be verified by looking at Figure 6.2. Note that the original matrix A is asymptotically stable, yet its ϵ-pseudospectral abscissa $\alpha_\epsilon(0) = -0.8$ is larger than $\alpha_\epsilon(x_*)$. Hence, optimizing $\alpha_\epsilon(x)$ over x yields a system that is more robustly stable compared to the original system. The subspace framework terminates after 3 subspace iterations, and the iterates generated are listed in Table 6.4. We also report the runtimes of the subspace framework in Table 6.5. Once again the computation of $\alpha_\epsilon(x)$ required several times to form and expand the subspaces dominate the runtime, whereas the minimization of the reduced pseudospectral abscissa functions is computationally much cheaper. The last column of Table 6.5 reports the total runtime in seconds for the direct minimization of $\alpha_\epsilon(x)$ (i.e., via `eigopt`), where we employ the subspace framework from [27] to compute the objective $\alpha_\epsilon(x)$; the direct minimization of $\alpha_\epsilon(x)$ takes about 8 times more computational time compared to that for the proposed subspace framework.

HF1 *($n = 130$, $m=1$, $p = 2$).* This example concerns the stabilization of $A+BKC$ with respect to $K \in \mathbb{R}^{1\times 2}$ for given $A \in \mathbb{R}^{130\times 130}, B \in \mathbb{R}^{130}, C \in \mathbb{R}^{2\times 130}$. We minimize $\alpha_\epsilon(x)$ using the subspace framework for $\epsilon = 0.2$ and $A(x) = A + x_1 BC(1,:) + x_2\, BC(2,:)$ over $x \in [-1, 1] \times [-1, 1]$, where $C(j,:)$ denotes the jth row of C. The subspace framework terminates after two subspace iterations with the optimal value of x as $x_* = (-0.36364, -0.26189)$, and the corresponding minimal value of the ϵ-

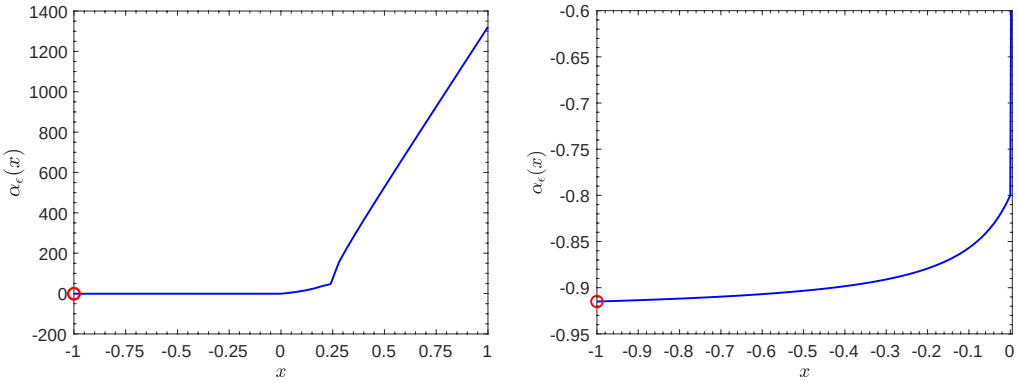

FIG. 6.2. *The plots of $\alpha_\epsilon(x)$ as a function of x for the NN18 example in the* COMPl_e*ib collection. The right-hand plot is a zoomed version of the left-hand plot focusing on $x \in [-1, 0.02]$.*

TABLE 6.5

Optimal values for $\alpha_\epsilon(x)$ over $x \in [-1, 1]$ for the NN18 example obtained using the subspace framework, as well as the runtimes for the subspace framework and direct minimization. The last four columns list in seconds the total runtime, the time for the reduced minimization problems, the time to compute a rightmost point in $\Lambda_\epsilon(x)$ several times for the subspace framework, and the total time for the direct minimization of $\alpha_\epsilon(x)$ without the subspace framework for minimization.

x_*	$\alpha_\epsilon(x_*)$	$\alpha_\epsilon(0)$	time	red	psa	direct
-1	-0.9149600	-0.8	22.9	3.9	18.2	182.9

pseudospectral abscissa as $\alpha_\epsilon(x_*) = 0.1740919$. The accuracy of these computed optimal values can be verified from Figure 6.3. The runtimes in Table 6.6 again confirm that the total runtime is determined by the time required for the computation of $\alpha_\epsilon(x)$ at several x.

HF2D2 *($n = 3796$, $m = 2$, $p = 3$).* This is a large-scale example that arises from a modeling of 2D heat flow [29, Section 3]. A stabilizer $K \in \mathbb{R}^{2\times 3}$ is sought so that $A + BKC$ is asymptotically stable for given $A \in \mathbb{R}^{3796\times 3796}, B \in \mathbb{R}^{3796\times 2}, C \in \mathbb{R}^{3\times 3796}$. The original matrix A is unstable with spectral abscissa 0.2556862, and ϵ-pseudospectral abscissa for $\epsilon = 0.2$ equal to 0.4625511.

We express $A + BKC$ in the form $A(x) = A + \sum_{j=1}^{6} x_j A_j$, where $x_j = k_{1j}$, $x_{3+j} = k_{2j}$, and $A_j = B(:, 1)C(j, :)$, $A_{3+j} = B(:, 2)C(j, :)$ for $j = 1, 2, 3$, while $B(:, 1)$ and $B(:, 2)$ represent the first and second columns of B. We minimize $\alpha_\epsilon(x)$ for $\epsilon = 0.2$ with the constraints that $x_j \in [-1, 1]$ for $j = 1, \ldots, 6$. The subspace framework terminates after 4 subspace iterations with optimal x_* corresponding to the following matrix:

$$K_* = \begin{bmatrix} 1 & 0.33494 & 1 \\ 1 & -1 & 1 \end{bmatrix}.$$

It appears from Figure 6.4 that the computed global minimizer is accurate. The resulting matrix is asymptotically stable, as indeed $\alpha_\epsilon(x_*) = -0.4124020$. As depicted in Figure 6.5, whereas one of the components of $\Lambda_\epsilon(0)$ is fully on the right-hand side of the complex plane, the rightmost component of $\Lambda_\epsilon(x_*)$ is fully contained in the open left-half of the complex plane. In terms of the runtime, now the reduced optimization problems take more time than the time to compute $\alpha_\epsilon(x)$ as reported in Table 6.7. Yet, the overall runtime for the subspace framework is quite reasonable considering

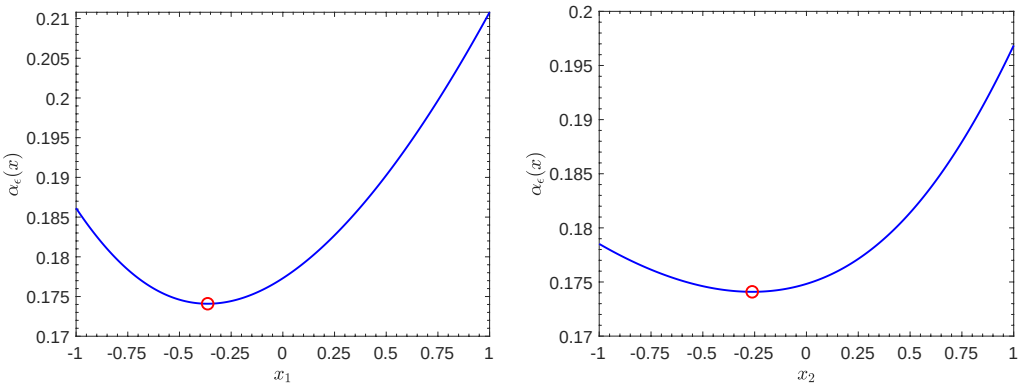

FIG. 6.3. *The plots of $\alpha_\epsilon(x)$ for $\epsilon = 0.2$ as a function of x for the HF1 example from the* COMPl_e*ib collection near its computed global minimizer $x_* = (x_{*,1}, x_{*,2})$. (Left) The plot of $x_1 \mapsto \alpha_\epsilon(x_1, x_{*,2})$. (Right) The plot of $x_2 \mapsto \alpha_\epsilon(x_{*,1}, x_2)$. The circles mark $(x_{*,1}, \alpha_\epsilon(x_*))$ and $(x_{*,2}, \alpha_\epsilon(x_*))$ on the left and on the right, respectively.*

TABLE 6.6
Optimal values for $\alpha_\epsilon(x)$ over $x \in [-1, 1] \times [-1, 1]$ for the HF1 example obtained using the subspace framework, and the runtimes for the subspace framework. The last three columns list the runtimes for the subspace framework in seconds as in Table 6.5.

x_*	$\alpha_\epsilon(x_*)$	$\alpha_\epsilon(0)$	time	red	psa
$(-0.36364, -0.26189)$	0.1740919	0.1810205	2.2	0.3	1.8

the system at hand is of order 3796, and there are several optimization parameters. Once again the time required for the direct minimization of $\alpha_\epsilon(x)$ (where again the objective $\alpha_\epsilon(x)$ is computed via the subspace framework in [27]) is significantly more than the overall runtime for the subspace framework.

7. Software. A Matlab implementation of the proposed subspace framework, that is Algorithm 2.1, is publicly available at https://zenodo.org/record/6992092. This implementation makes use of all of the implementation details described in Section 6.1. The numerical results on the benchmark examples in Section 6.3 can be reproduced by running the script `demo_on_benchmarks`.

8. Concluding Remarks. Minimization of the ϵ-pseudospectral abscissa of a matrix dependent on parameters for a prescribed $\epsilon > 0$ is motivated by robust stability and transient behavior considerations for the associated linear control system, as well as stabilization problems such as the stabilization by static output feedback. Here, we have proposed a subspace framework to minimize the ϵ-pseudospectral abscissa of a large matrix-valued function dependent on parameters analytically aiming at large-scale nature of the matrix-valued function. The large-scale matrix-valued functions are restricted to small subspaces, and the ϵ-pseudospectral abscissa of the resulting reduced small-scale matrix-valued functions is minimized. The subspaces are gradually expanded so as to attain Hermite interpolation properties between the ϵ-pseudospectral abscissa of the original and reduced matrix-valued functions at the minimizers of the reduced problems. We have proven the global convergence of the subspace framework in the infinite dimensional setting, that is the convergence of the optimal values of the reduced problems to the globally smallest value of the

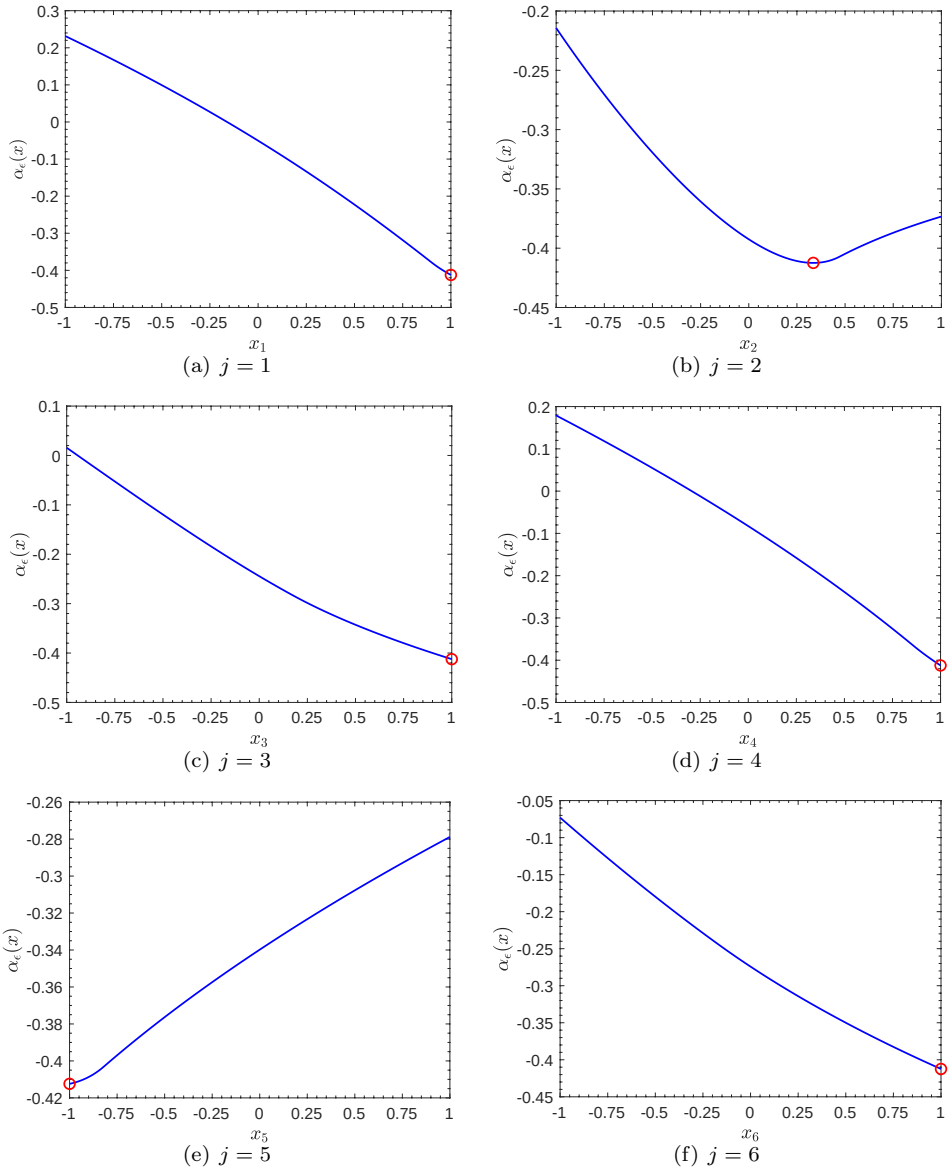

FIG. 6.4. *The plots of $\alpha_\epsilon(x)$ for $\epsilon = 0.2$ as a function of x for the HF2D2 example near the computed global minimizer $x_* = (x_{*,1}, x_{*,2}, x_{*,3}, x_{*,4}, x_{*,5}, x_{*,6})$. Each plot corresponds to the graph of $x_j \mapsto \alpha_\epsilon(x_{*,j}[x_j])$, where $x_{*,j}[x_j]$ is equal to x_* except its jth component is x_j. The circle marks $(x_{*,j}, \alpha_\epsilon(x_*))$.*

TABLE 6.7

Optimal values of $\alpha_\epsilon(x)$ over $x \in [-1, 1]^6$ for the HF2D2 example by the subspace framework, and the runtimes for the subspace framework, as well as direct minimization. The last four columns are runtimes in seconds as in Table 6.5.

$\alpha_\epsilon(x_*)$	$\alpha_\epsilon(0)$	time	red	psa	direct
−0.4124020	0.4625511	36.2	25.5	9.6	211.5

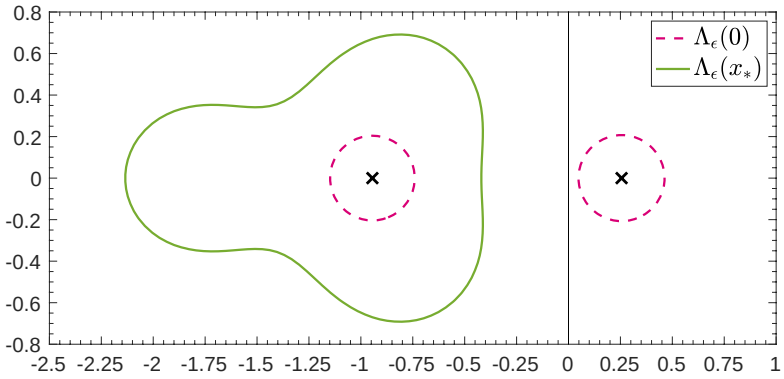

FIG. 6.5. *The plots of of the boundaries of the rightmost components of $\Lambda_\epsilon(0)$ and $\Lambda_\epsilon(x_*)$ for the HF2D2 example. The crosses mark the rightmost two eigenvalues of A, while the vertical line represents the imaginary axis.*

ϵ-pseudospectral abscissa of the original matrix-valued function, under mild assumptions. Additionally, we have shown that the rate of convergence of the minimizers of the reduced problems to the global minimizer of the original problem is superlinear, again under mild assumptions. The validity of these theoretical findings in practice is confirmed on synthetic and benchmark examples. The proposed framework makes it feasible to minimize the pseudospectral abscissa of matrix-valued functions of size on the order of thousands, such as the NN18 and HF2D2 examples from the *COMPl$_e$ib* collection, in a short time.

Some of the ingredients in an actual implementation of the overall subspace framework are locally convergent. Specifically, the subspace framework requires the rightmost point of the ϵ-pseudospectrum of the original matrix-valued function at several parameter values, and if the matrix-valued function is really large, then it appears for this purpose one has to rely on a subspace framework such as the one proposed in [27] that converges locally. Additionally, if the minimization is over several parameters, it appears that the reduced minimization problems must be solved by employing a locally convergent optimization algorithm such as those based on Newton's method, e.g., "GRANSO" [18]. There is no quick remedy for these local convergence issues. However, on several benchmark examples considered here, the subspace framework indeed converges globally.

The reduced minimization problems require computing the ϵ-pseudospectral abscissa of rectangular pencils as in (2.2). We solve at the moment such pseudospectral abscissa problems using an extension of the criss-cross algorithm [11] to rectangular pencils [27, Section 5.2]. One major challenge in the rectangular setting is that one has to start from a point in the rightmost connected component of the associated rectangular ϵ-pseudospectrum. We have outlined some ideas to overcome this challenge in Section 6.1.3. Yet, locating the rightmost point in the ϵ-pseudospectrum, and computing the ϵ-pseudospectral abscissa for a rectangular pencil appear to be far from settled. This is a problem worth investigating in detail.

Acknowledgements. The authors are grateful to two anonymous referees, and the associate editor handling the manuscript for their invaluable feedback.

Appendix A. Proof of Theorem 2.4. As $\widetilde{x}$ is nondegenerate, there is a unique optimal z for the maximization problem in (2.4) with $x = \widetilde{x}$, which we denote

by $z(\widetilde{x})$. Moreover, the singular value $\sigma(\widetilde{x}, z(\widetilde{x}))$ is simple, so $\sigma(x, z)$ is real-analytic at $(x, z) = (\widetilde{x}, z(\widetilde{x}))$. By the first order optimality conditions applied to the constrained optimization problem in (2.4) with $x = \widetilde{x}$, there is $\mu(\widetilde{x})$ such that

$$\mathcal{L}_{z_1}(\widetilde{x}, z(\widetilde{x}), \mu(\widetilde{x})) = 1 - \mu(\widetilde{x})\sigma_{z_1}(\widetilde{x}, z(\widetilde{x})) = 0.$$

Clearly, $\mu(\widetilde{x})$ is uniquely defined by $\mu(\widetilde{x}) = 1/\sigma_{z_1}(\widetilde{x}, z(\widetilde{x}))$, and $\mu(\widetilde{x}) \neq 0$, which in turn leads to $\sigma(\widetilde{x}, z(\widetilde{x})) = \epsilon$ by the complementarity conditions.

By continuity, the singular value function $\sigma(x, z)$ remains simple and real-analytic in a neighborhood of $(\widetilde{x}, z(\widetilde{x}))$. Hence, from a standard sensitivity analysis for optimization problems, there is an open neighborhood X of $\widetilde{x}$ such that for every $x \in X$ the optimal z as well as corresponding Lagrange multiplier μ for the maximization problem in (2.4) are unique, and the unique optimizer $y(x) = (z(x), \mu(x))$ is differentiable at every $x \in X$. From the first order optimality conditions, we have

$$\nabla_y \mathcal{L}(x, y(x)) = 0 \quad \forall x \in X. \tag{A.1}$$

Moreover, since $\widetilde{x}$ is nondegenerate, the matrix $\nabla_{yy}\mathcal{L}(\widetilde{x}, y(\widetilde{x}))$ is nonsingular. Hence, the analytic implicit function theorem guarantees the existence of a unique real-analytic function $w : U \to \mathbb{R}$ satisfying $w(\widetilde{x}) = y(\widetilde{x})$ and $\nabla_y \mathcal{L}(x, w(x)) = 0\ \forall x \in U$ in an open neighborhood $U \subseteq X \subseteq \mathbb{R}^d$ of $\widetilde{x}$. The function $w(x)$ is also the unique differentiable function satisfying $\nabla_y \mathcal{L}(x, w(x)) = 0\ \forall x \in U$ and $w(\widetilde{x}) = y(\widetilde{x})$. As a result, we have $w(x) = y(x)$ for all $x \in U$. This shows that $x \in U \mapsto \alpha_\epsilon(x) = z_1(x)$ is real-analytic at $x = \widetilde{x}$.

Using (A.1) as well as the analytical formulas for the derivatives of singular value functions (see, e.g., [33, Section 3.3]), we deduce

$$0 = \mathcal{L}_{z_1}(\widetilde{x}, z(\widetilde{x}), \mu(\widetilde{x})) = 1 - \mu(\widetilde{x}) \operatorname{Re}\left(u^* \frac{\partial}{\partial z_1}[A(x, z)]\Big|_{\widetilde{x}, z(\widetilde{x})} v \right) = 1 + \mu(\widetilde{x}) \operatorname{Re}(u^* v)$$

and

$$0 = \mathcal{L}_{z_2}(\widetilde{x}, z(\widetilde{x}), \mu(\widetilde{x})) = -\mu(\widetilde{x}) \operatorname{Re}\left(u^* \frac{\partial}{\partial z_2}[A(x, z)]\Big|_{\widetilde{x}, z(\widetilde{x})} v \right) = -\mu(\widetilde{x}) \operatorname{Im}(u^* v),$$

where u, v consist of a pair of consistent unit left and right singular vectors corresponding to $\sigma(\widetilde{x}, z(\widetilde{x}))$. The first equation implies $\operatorname{Re}(u^* v) \neq 0$. Furthermore, as $\mu(\widetilde{x}) \neq 0$, the second equation gives rise to $\operatorname{Im}(u^* v) = 0$, that is $u^* v$ is real and so $u^* v = \operatorname{Re}(u^* v)$. It follows from the first equation that

$$\mu(\widetilde{x}) \;=\; -\frac{1}{u^* v}.$$

From a classical sensitivity analysis result, in particular the fact that the gradient with respect to the parameters of the optimal value of an optimization problem dependent on parameters is equal to the gradient of the Lagrangian with respect to the parameters, we deduce

$$\nabla \alpha_\epsilon(x) = \nabla_x \mathcal{L}(x, y(x)) \tag{A.2}$$

for every $x \in U$. More specifically, denoting the jth component of x with x_j, we have

$$\frac{\partial}{\partial x_j}[\alpha_\epsilon(x)]\Big|_{\widetilde{x}} = -\mu(x) \frac{\partial}{\partial x_j}[\sigma(x, z)]\Big|_{\widetilde{x}, z(\widetilde{x})}.$$

Again, using the analytical formula for the singular value function, and $\mu(\widetilde{x}) = -1/(u^*v)$, which is real and nonzero, we obtain (2.6).

For the second derivatives, we differentiate (A.2) by applying the chain rule to obtain

$$\nabla^2\alpha_\epsilon(x) \;=\; \nabla^2_{xx}\mathcal{L}(x,y(x)) + \nabla^2_{xy}\mathcal{L}(x,y(x))\cdot y'(x)\;, \tag{A.3}$$

where $y'(\cdot)$ denotes the $3\times d$ Jacobian matrix of y with respect to x. Differentiating (A.1) with respect to x yields

$$\nabla^2_{yx}\mathcal{L}(x,y(x)) + \nabla^2_{yy}\mathcal{L}(x,y(x))\cdot y'(x) \;=\; 0\;.$$

In particular, as $\widetilde{x}$ is nondegenerate, $\nabla^2_{yy}\mathcal{L}(\widetilde{x},y(\widetilde{x}))$ is invertible, so

$$\begin{aligned} y'(\widetilde{x}) \;&=\; -[\nabla^2_{yy}\mathcal{L}(\widetilde{x},y(\widetilde{x}))]^{-1}\nabla^2_{yx}\mathcal{L}(\widetilde{x},y(\widetilde{x})) \\ &=\; -[\nabla^2_{yy}\mathcal{L}(\widetilde{x},y(\widetilde{x}))]^{-1}[\nabla^2_{xy}\mathcal{L}(\widetilde{x},y(\widetilde{x}))]^T. \end{aligned} \tag{A.4}$$

Substituting the right-hand side of (A.4) for $y'(\widetilde{x})$ in (A.3) with $x=\widetilde{x}$, we obtain (2.7). □

REFERENCES

[1] N. Aliyev, P. Benner, E. Mengi, P. Schwerdtner, and M. Voigt, *Large-scale computation of $\mathcal{L}_\infty$-norms by a greedy subspace method*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 1496–1516.

[2] N. Aliyev, P. Benner, E. Mengi, and M. Voigt, *A subspace framework for $\mathcal{H}_\infty$-norm minimization*, SIAM J. Matrix Anal. Appl., 41 (2020), pp. 928–956.

[3] P. Apkarian and D. Noll, *Controller design via nonsmooth multidirectional search*, SIAM J. Control Optim., 44 (2006), pp. 1923–1949.

[4] P. Apkarian and D. Noll, *Nonsmooth H_∞ synthesis*, IEEE Trans. Autom. Control, 51 (2006), pp. 71–86.

[5] P. Apkarian, D. Noll, and O. Prot, *A trust region spectral bundle method for nonconvex eigenvalue optimization*, SIAM J. Optim., 19 (2008), pp. 281–306.

[6] D. Arzelier, G. Deaconu, S. Gumussoy, and D. Henrion, *H_2 for HIFOO*, in Int. Conference on Control and Optimization with Industrial Applications, Ankara, Turkey, August 2011.

[7] P. Benner and T. Mitchell, *Extended and improved criss-cross algorithms for computing the spectral value set abscissa and radius*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 1325–1352.

[8] B. Bernhardsson, A. Rantzer, and L. Qiu, *Real perturbation values and real quadratic forms in a complex vector space*, Linear Algebra and its Applications, 270 (1998), pp. 131–154.

[9] V. Blondel, M. Gevares, and A. Lindquist, *Survey on the state of systems and control*, European J. Control, 1 (1995), pp. 5–23.

[10] V. Blondel and J. N. Tsitsiklis, *NP-hardness of some linear control design problems*, SIAM J. Control Optim., 35 (1997), pp. 2118–2127.

[11] J. Burke, A. S. Lewis, and M. L. Overton, *Optimization and pseudospectra, with applications to robust stability*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 80–104, https://doi.org/10.1007/s10589-017-9898-5.

[12] J. V. Burke, D. Henrion, A. S. Lewis, and M. L. Overton, *HIFOO – A MATLAB package for fixed-order controller design and H_∞ optimization*, in Proc. 5th IFAC Syposium on Robust Control Design, Toulouse, France, Jul. 2006.

[13] J. V. Burke, A. S. Lewis, and M. L. Overton, *Optimal stability and eigenvalue multiplicity*, Found. Comput. Math., 1 (2001), pp. 205–225.

[14] J. V. Burke, A. S. Lewis, and M. L. Overton, *Optimizing matrix stability*, Proc. Am. Math. Soc., 129 (2001), pp. 1635–1642.

[15] J. V. Burke, A. S. Lewis, and M. L. Overton, *Two numerical methods for optimizing matrix stability*, Linear Algebra Appl., 351-352 (2002), pp. 117–145.

[16] J. V. Burke, A. S. Lewis, and M. L. Overton, *A nonsmooth, nonconvex optimization approach to robust stabilization by static output feedback and low-order controllers*, in Proc. 4th IFAC Syposium on Robust Control Design, Milan, Italy, June 2003.

[17] J. V. Burke, A. S. Lewis, and M. L. Overton, *Robust stability and a criss-cross algorithm for pseudospectra*, IMA J. Numer. Anal., 23 (2003), pp. 359–375.

[18] F. E. Curtis, T. Mitchell, and M. L. Overton, *A BFGS-SQP method for nonsmooth, nonconvex, constrained optimization and its evaluation using relative minimization profiles*, Optim. Method Softw., 32 (2017), pp. 148–181.

[19] N. Guglielmi and G. Lubich, *Low-rank dynamics for computing extremal points of real pseudospectra*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 40–66.

[20] N. Guglielmi and M. L. Overton, *Fast algorithms for the approximation of the pseudospectral abscissa and pseudospectral radius of a matrix*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1166–1192.

[21] S. Gumussoy and W. Michiels, *Computing $\mathcal{H}_\infty$ norms of time-delay systems*, in Proc. Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, Shanghai, China, Dec. 2009, pp. 263–268.

[22] M. Gurbuzbalaban and M. L. Overton, *Some regularity results for the pseudospectral abscissa and pseudospectral radius of a matrix*, SIAM J. Optim., 22 (2012), pp. 281–285.

[23] F. Kangal, K. Meerbergen, E. Mengi, and W. Michiels, *A subspace method for large scale eigenvalue optimization*, SIAM J. Matrix Anal. Appl., 39 (2017), pp. 48–82.

[24] T. Kato, *Perturbation Theory for Linear Operators*, Springer, Berlin, 1995.

[25] H. O. Kreiss, *Über die stabilitätsdefinition für differenzengleichungen die partielle differenzialgleichun- gen approximieren*, BIT, 2 (2012), pp. 153–181.

[26] D. Kressner, D. Lu, and B. Vandereycken, *Subspace acceleration for the crawford number and related eigenvalue optimization problems*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 961–982.

[27] D. Kressner and B. Vandereycken, *Subspace methods for computing the pseudospectral abscissa and stability radius*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 292–313.

[28] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, PA, 1998.

[29] F. Leibfritz, COMP*leib:* CO*nstraint* M*atrix-optimization* P*roblem* lib*rary – a collection of test examples for nonlinear semidefinite programs, control system design and related problems*, tech. report, Department of Mathematics, University of Trier, 2004.

[30] A. S. Lewis and J. P. Pang, *Variational analysis of pseudospectra*, SIAM J. Optim., 19 (2008), pp. 1048–1072.

[31] D. Lu and B. Vandereycken, *Criss-cross type algorithms for computing the real pseudospectral abscissa*, SIAM J. Matrix Anal. Appl., 38 (2017), pp. 891–923.

[32] E. Mengi, *Large-scale and global maximization of the distance to instability.*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 1776–1809.

[33] E. Mengi, E. A. Yildirim, and M. Kiliç, *Numerical optimization of eigenvalues of Hermitian matrix functions*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 699–724, https://doi.org/10.1137/130933472, http://dx.doi.org/10.1137/130933472.

[34] A. Nemirovskii, *Several NP-hard problems arising in robust stability analysis*, Math. Control Signals Syst., 6 (1993), pp. 99–105.

[35] M. L. Overton, *HANSO: a hybrid algorithm for nonsmooth optimization*, 2009. http://cs.nyu.edu/overton/software/hanso.

[36] M. W. Rostami, *New algorithms for computing the real structured pseudospectral abscissa and the rel stability radius of large and sparse matrices*, SIAM J. Sci. Comput., 37 (2015), pp. S447–S471.

[37] L. N. Trefethen, *Computation of pseudospectra*, Acta Numer., 8 (1999), pp. 247–295.

[38] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princetion University Press, Princeton, NJ, 2005.